MODELING AND FORECASTING OF ELECTRICITY CONSUMPTION USING ARMA AND EXPONENTIAL SMOOTHING:

A Case Study of Gazaoua Plant

BY

Yakubu HUSSAIN

P16PSST8041

DEPARTMENT OF STATISTICS,
AHMADU BELLO UNIVERSITY,
ZARIA, NIGERIA

MARCH, 2018

MODELING AND FORECASTING OF ELECTRICITY CONSUMPTION USING ARMA AND EXPONENTIAL SMOOTHING:

A Case Study of Gazaoua Plant

BY

Yakubu HUSSAIN

P16PSST8041

A DISSERTATION SUBMITTED TO THE SCHOOL OF POSTGRADUATE STUDIES, AHMADU BELLO UNIVERSITY, ZARIA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE IN STATISTICS

MARCH, 2018

DECLARATION

I hereby declare that the work in this dissertation entitled ‘‘Modeling and Forecasting of Electricity Consumption using Autoregressive Moving Average and Exponential Smoothing Methods’’at Gazaoua Power Plant, has been performed by me in the Department of Statistics of Ahmadu Bello University Zaria (A.B.U), under the supervision of Dr. A. Yahaya and Dr. S. I. S. Doguwa. The information derived from the literature has been duly acknowledged in the text and a list of references provided. No part of thisdissertation was previously presented for another degree or diploma at any other Institution.

_____ _____

Name of studentSignatureDate

CERTIFICATION

This Dissertation entitled “Modeling and Forecasting of Electricity Consumption using Autoregressive Moving Average and Exponential Smoothing Methods”at Gazaoua Power Plantby HUSSAIN Yakubu, meets the regulations governing the award of the degree of Master of Science of Ahmadu Bello University, Zaria and is approved for its contribution to knowledge and literary presentation.

_____ _____

Dr. A. Yahaya
Chairman, Supervisory Committee Date

_____ _____

Dr. S.I.S. Doguwa Date Member,
Supervisory Committee

_____ _____

Dr. O. E. Olubusoye
External Examiner Date

_____ _____

Dr. H. G. Dikko Date
Head of Department

_____ _____

Prof. Sadiq Z. Abubakar Date
Dean, School of Postgraduate Studies

DEDICATION

This dissertation is dedicated to my late Parents, Alhaji Hussain Aliyu and Hajiya Maryam Aliyu.

ACKNOWLEDGEMENTS

I wish to express my sincere and heartfelt gratitude to my research supervisors, Dr. A. Yahaya and Dr. S.I.S. Doguwa for their guidance, corrections, comments and constructive suggestions throughout the duration of the research work. Their commitment and patience in going through the work tirelessly and thoroughly will forever be remembered and I will forever remain grateful for their contribution towards successful completion of this research work. My sincere regards and appreciation is extended to my colleagues in the Faculty of Computing and Mathematical Sciences, Kano University of Science and Technology, Wudil such as Dr. Ibrahim Abdullahi, Dr. M.L. Kaurangini, Dr. Abubakar Musa, Dr. Abdulhadi Aminu, Dr. Salisu M. Tudunkaya, Dr. Salisu Usaini, Mal. Sani Salihu Abubakar, Mal. Lawal Adamu Ismail, Mal. Munkaila Danjuma, Mal. Auwalu Ibrahim, Mal. Hassan Umar, Mal. Musa Uba, Mal. Mukhtar Aliyu, Mal. Yusuf I. Gwanda, Mal. Miftahu Zubairu Ringim, Mal. Musa Yusuf, Mal. Kabiru Dalhatu, Mal. Alasan, Mal. Sadiq, Mal. Mustapha Abdulkadir, and Mal. Aliyu Yusuf Ahmed.

My special regards go to Engr. Liman Panshekara, Engr. Abdullahi Panshekara (Abdul), Engr. Ibrahim Panshekara and Sagir Hukuma. My sincere and honest gratitude go to my elder brother Alhaji Jibrin Hussain for playing the role of a father right from childhood. His financial and moral support is highly appreciated. I appreciate the support and understanding exhibited during the research by my able sister Hajia Binta Hussain Aliyu. I would also like to salute my beloved wife Aisha Adam for support and advices given. Lastly, I wish to pray for Aljannaat Alfirdaus for my late parents: Alhaji Hussain Aliyu and Hajiya Maryam Aliyu. May their sins be erased and have eternal rest.

ABSTRACT

The various available literature on modeling electricity consumption use time series models such as Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), Vector Auto-regression (VAR) and so on. In this research, Autoregressive Moving Average (ARMA)and Exponential Smoothing processes are used to fit the monthly electricity consumption (mega watt hour) from January, 2011 to Dec, 2015 (a case study of Gazaoua power plant). Augmented Dickey Fuller (ADF) test was performed to check for presence of unit root. The work, based on the Autocorrelation function (ACF) and Partial Autocorrelation function (PACF) examines the forecast ability of varieties of Autoregressive Moving Average (ARMA) and Exponential Smoothing. The result shows that seasonal Autoregressive Moving Average of first order and Holt-winters Multiplicative satisfied the best model criteria. However, it is further observed based on loss function that Holt-winters Multiplicative emerged as the best forecasting method for the data set.

TABLE OF CONTENTS

COVER PAGE.....i
FLY LEAF.....ii
TITLE PAGE.....iii
DECLARATION.....iv
CERTIFICATION.....v
DEDICATION.....vi
ACKNOWLEDGEMENTS.....vii
ABSTRACT.....viii
TABLE OF CONTENTS.....ix
LIST OF ABBREVIATIONTS.....xiii

CHAPTER ONE: INTRODUCTION

1.1 Background to the Study.....1
1.2 Problem Statement.....4
1.3 Aim and Objectives of the Study.....4
1.4 Significance of the Study.....5
1.5 Scope and Limitation of the Study.....5

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction.....6
2.2 The Econometric Modeling Approach.....6
2.3 The Log Linear Models.....7

CHAPTER THREE: METHODOLOGY

3.1 Time Series Model 9
3.2 Method of Data Collection 12
3.3 Method of Data Analysis 12
3.4 Statistical Parameters 13
3.4.1 Mean 13
3.4.2 Variance 13
3.4.3 Skewness 14
3.4.4 Kurtosis 14
3.4.5 Coefficient of Serial Correlation 14
3.5 Jarque-Bera 14
3.6 Identification of Distribution 15
3.7 Autocorrelation Function 15
3.7.1 Expectation Function 15
3.7.2 Autocovariance Function 15
3.8 Partial Autocorrelation Function 17
3.9 Stationarity 17
3.9.1 Moments of Stationary Process 18
3.10 Unit Root Test 20
3.11 Seasonality 21
3.12 Univariate Time Series Process 22
3.12.1 Stochastic Process 22
3.12.2 Autoregressive Process 22
3.12.3 Moving Average Process 24
3.13 ARMA Process 26
3.13.1 Missing Data 26
3.13.2 Outlier 26
3.13.3 Stationarity 26
3.13.4 Normal Distribution 27

3.14 Seasonal ARMA Process 28
3.15 Akaike's Information Criterion 30
3.15.1 Corrected Akaike's Information Criterion 30
3.16 Model Parsimony 31
3.17 Model Building Process 31
3.17.1 Model Identification 31
3.17.2 Parameter Estimation 32
3.17.3 Model Verification 32
3.18 Comparison of Competing Models and Forecast Evaluation..... 34
3.19 Static and Dynamic Forecasting..... 35
3.20 Model Based Forecasting 35
3.21 Exponential Smoothing 37
3.21.1 Single Exponential Smoothing 37
3.21.2 Double Exponential Smoothing 37
3.21.3 Holt-Winters Multiplicative 37
3.21.4 Holt-Winters Additive 38

CHAPTER FOUR: RESULTS AND DISCUSSION

4.1 Introduction…..... 39
4.2 Data Analysis …..... 39
4.2.1 Model Identification 42
4.2.2 Estimation of Parameter 44
4.2.3 Diagnostic Checking 45
4.3 Exponential Smoothing 47
4.4 Model Comparison and Forecast Evaluation … 48
4.5 Contribution to Knowledge… 48

CHAPTER FIVE: SUMMARY, CONCLUSION &RECOMMENDATIONS

5.1 Summary.....49
5.2 Conclusion 49
5.3 Recommendations.....49
APPENDIX 50
REFERENCES 52

LIST OF ABBREVIATIONS

1.	**ACF**	Autocorrelation function
2.	**ADF**	Augmented Dicky Fuller
3.	**AIC**	Akaike Information Criterion
4.	**AICc**	CorrectedAkaike Information Criterion
5.	**AR**	Autoregressive
6.	**ARCH**	Autoregressive Conditional Heteroskedasticity
7.	**ARMA**	Autoregressive Moving Average
8.	**BG**	Breusch-Godfrey
9.	**Corr**	Correlation
10.	**Cov**	Covariance
11.	**d**	Differencing term
12.	**DF**	Dicky Fuller
13.	**JB**	Jarque-Bera
14.	**K**	Kurtosis
15.	**MA**	Moving Average
16.	**MAE**	Mean Absolute Error
17.	**p**	Autoregressive term
18.	**PACF**	Patial Autocorrelation Function
19.	**q**	Moving Average term
20.	**Q-Q**	Quantile-Quantile
21.	**RMSE**	Root Mean Square Error
22.	**RSS**	Residual Sum of Squares
23.	**SARMA**	Seasonal Autoregressive Moving Average
24.	μ	Mean

25.	S^2	Variance
26.	**Var**	Variance
27.	**WN**	White Noise
28.	**X-bar**	Mean

CHAPTER ONE

INTRODUCTION

1.1 Background to the Study

In the early seventies, there has been an increase in the number of research works on energy and power demand(Pindyck, 1979). The debate between economists and engineers over the years informed development of methodologies for modeling energy need and thus helped in developing variety of models for analysing and forecasting energy demand(Wirl *et al*., 1990). Ryan *et al*., (2009) have the argued that computing data availability on power and the training of energy analysts usually developed with time and as a result demand modeling has advanced to a greater extent than the early researches in modeling of energy demand. Energy is a developed demand rather than a need for its own sake; it is developed from the demand of the end use services that utilize energy resources with the capital stock that uses energy resources to provide these end-use services such as lighting, heating etc. (Hartman *et al*., 1979). Therefore, analysis of energy demand should be implicitly or explicitly, admit the fact that energy consuming appliances and energy resources are combined in different ways to have these services in place.

The behaviour of energy demand can be summarized in three steps. Firstly, the energy user decides whether to buy energy consuming and durable appliances that will allow for a particular service. Secondly, the user has a choice about the technical and economic characteristics of the appliances. Thirdly, the consumer's choice of the frequency of use and intensity of that appliance will influence the level of use or demand(Hartman *et al*., 1979). The capital stock and its characteristics are generally assumed to be stilled in the short run. Therefore the energy demand behaviour in the short run might be different from that in the long run. For instance, the households' decision about whether to buy a new residential appliance depends on household income, the climate in which he lives, the cost of purchasing

&operating cost for the appliance and the general socioeconomic trends patterns that affect the popularity of such appliances. The choice of economic and technological characteristics of appliances depends upon the comparison of capital and operating costs, reliability, size and efficiency of alternatives. Moreover, the climate or the region where the appliance is used might affect the choice of fuel and other characteristics of the appliances; once the decision about the residential appliance has been made, the capital stock is fixed in the short run. Therefore, the capital utilization of these appliances depends upon the cost of the fuel used by the appliance, income and the other characteristics of the household.

Hartman, *et al*., (1979) argued that an energy demand model should analyse three sets of decision stated above by taking into account the characteristics of the energy user, the technical and economic characteristics of the energy source and the capital stock, and the characteristics of the environment that the capital stock is used. As the policy implications of energy demand models are important, Hartman (1979) furthermore states that the variables subject to policy control or that might affect or guide the energy user decisions should be included. However, there are number of different approaches to model energy demand. According to Ryan *et al*., (2009), there is no single 'right' approach to modeling energy demand, the modeling strategy might differ according to a range of conditions and here are different approaches being categorized into three main groups: i) end-use modeling; ii) input-output modeling; iii) econometric modeling.

One of the earliest researches based on the end-use or engineering-economy modeling approach was popularly known as the bottom up approach(Chateau *et al*., 1978). This approach is based on predicting the energy demand in different sectors or industries using the relationship between output and energy use. The data required for end-use modeling approach is obtained through energy surveys, technical studies, and energy audits and focuses on dividing the sectoral demand into homogeneous parts, so that the energy demand for each

part can beeasily related to the technical and economic factors - the key factors that determine the energy demand for each sector. According to Bhattacharyya and Timilsina (2009), the general process of end-use modeling is summarized as follows:

- Total energy demand is disaggregated into homogenous end use categories;
- The evaluation process of social, economic, and technological factors in order to identify the interrelationships and long term development;
- The determinants are organized into a hierarchical structure;
- The mathematical formulization of the hierarchical structure according to the identified relations;
- A snap-shot view of reference year;
- Different scenarios are designed for the future based on a variety of assumptions about the determinants; and
- Forecasting takes place according to scenarios and the mathematical relationship between the determinants.

Several models have been developed considering the level of disaggregation, technology selection, technology representation, model target and the level of macroeconomic integration (Worrel *et al*., 2004). In reviewing a range of energy demand models for policy formulation, Bhattacharyya *et al.*(2009) highlights that, most end-use demand models do not rely on neo-classical economic theory. Moreover, they do not focus on history but recent structural changes and technological developments, which are arguably the main strength of this approach. These models also search for the optimal level of aggregation of sectors by categories that generate satisfactory homogenous consumer groups; for example the rural-urban divide. However, the level of disaggregation in the end-use approach is often not supported by available data in most of the cases; therefore, data limitation is seen as a major weakness of this approach(Pesaran *et al.,* 1998). Another weakness of this approach is that

the accounting type end-use models are unable to identify the price induced effects(Bhattacharyya *et al.,* 2009). Meanwhlie, price effects are important for policy makers! Wassily Leontief (1920) developed the input-output approach. This systematically quantifies the interrelationships between ranges of sectors in a complex economic system and based on a fully determined general equilibrium model (Arbex and Perobelli, 2010). This analyses theprocess in which inputs from one industry produce output for consumption or input for another industry. The application of this approach to energy demand enables the estimation of the direct energy demand as well as indirect energy demand via inter-industry transactions (Bhattacharyya *et al.*, 2009).

1.2 Problem Statement

The rapid growth of the population affects the consumption of electricity. When population is fast growing there is need to predict future demands of electricity supply so as to get prepared ahead thereby providing basic devices. Time series forecasting methods can be used to predict electricity consumption. These forecasting methods need to be inspected for their ability to providereliable forecast so as to enable the policy makers to make proper arrangement in order to meet up the need of the rapid growth of the population in future.

1.3 Aim and Objectives of the study:

The aim of this research work is to fit a time series forecasting modelto monthly electricity consumption (mwh).

The objectives of the study are as follows

1. Check for existence of unit root.
2. Investigate the behavioral change in the data using ARMA and Exponential Smoothing processes
3. Evaluate the accuracy of the models for forecasting capability.

1.4 Significance of the Study

It is hope that, the results obtained from this research work will be of great importance and relevance to energy and resource sectors.

1.5 Scope and Limitation

The research is based on electricity consumption (mwh) at Gazaoua power plant. The data used for the research was secondary and on Monthly basis for a period of five years (January, 2011 to December, 2015). The data was obtained from Kumbotso Sub-Region Transmission Power Plant, Kano State. However, the researcher was constrained by time and finance.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter gives literature reviews on the works of different scholars in Time series analysis. Several research works have now used the forecasting method with time series data such as the electric power consumption.

2.2 The Econometric Modeling Approach

The modeling of energy demand in econometric is plan with the aim to analyse relationship between variables. The past relationship can be analysed based on the observed relationships, estimating the effect of changes of the independent variable(s) on the dependent variable. The modeling technique of econometric has been frequently used for analysing energy demand because, of the availability of observations. It can be applied with sufficiently long historical observations on energy consumption, and explanatory variables such as population, income, and prices. For the end-use and input-output modeling approaches, the main strategy is the homogenous grouping of consumers in order to model common characteristics of the energy demand of these homogenous consumer groups (industrial, residential, etc.)

Although this strategy is utilized by the econometric modeling approach, the main difference between this and the two other approaches is that the econometric modeling approach statistically estimates energy demand relationships; the end-use and input-output approaches rely on energy surveys and technical studies which are not always available. One of the reasons that the econometric approach is arguably more attractive than the other approaches is that the econometric approach has a strong theoretical background consistent with economic theory (in particular consumer and production theory). In the econometrics literature there are several functional forms which have been developed for energy demand modeling such as the trans-log model (most often applied to a demand system).

2.3 The Log Linear Models and Their Applications

The demand for energy is not a final demand; the energy demand is generated because of the demand for goods and services which needs energy in order to be utilized; such as heat, light, transport, etc. (Nordhaus, 1977). Therefore, the stock of appliances and its capacity usage are important factors that contribute to determining energy demand. This relationship according to Bohi *et al.*,(1984) can be shown as follows:

$$E_t = F(A_t, R_t) \tag{2.1}$$

where;

E_t = total demand for aggregated energy; A_t = stock of appliance for aggregated energy;

R_t = capacity usage rate of the appliances; and t = time period t

Various forecasting problems involve the use of time series observations. Time series can be used in forecasting of future events from current and past values of a phenomenon (Box and Jenkins, 1976). Time series plots can revealpatterns such as trends, random, seasonality, periods, unusual observations, and ora combination of patterns.

Saab *et al*., (2001) used ARIMA and AR(1) on monthly electricity consumption. The result recommended the AR (1).

Contreras *et al*., (2003) used ARIMA methods to estimate the next day electricity price. They propose two ARIMA models to predict hourly prices in the electricity markets of Spain and California, respectively.The Spanish model needs 5 hours to predict future prices, as opposed to the 2 hours needed by the Californian model. These differences may reflect different bidding structures and ownership.

Conejo *et al*., (2005) applied the wavelet transform and ARIMA models to predict electricity price. The functioning of the proposed technique,whichisbasedonthewavelet transformandARIMAmodels, outperformsthe direct use of ARIMA models.

Ediger (2007) used ARIMA model to predict electricity consumption in Turkey. The result showed an increase in the consumption at an annual average of 3.3%

Zhu *et al*., (2012) used ARIMA and VAR forecasting methods. Both the two methods were recommended.

Doguwa and Alade (2015)compared SARIMA and ARDL processes with the existing ARIMA process on external reserves in Nigeria. The results recommend SARIMA modelin three to six months forecast horizon, whereas ARDL model performs better in one to two months forecast horizon.

Musund and M'mukiira**(2016)** used ARIMA to model the GDP in Kenya from 1960 to 2012. ARIMA (2, 2, 2) was established as the best model under the AIC criterion.

In this study, ARMA and Exponential smoothing models are proposed using monthly electricity consumption (mwh).

CHAPTER THREE

METHODOLOGY

3.1 Time Series Model

The procedure employed in analyzing data is presented in this section. Time series linear models have been proposed and compared for forecasting electricity consumption (mwh). Moreover, the methods for model identification in forecasting ability have also been discussed. Time series can be defined as a chronological series of observations (Montgomery et al., 2008). The time can be a discrete, or a continuous function. Time series data can be deterministic, random, or a combination of the two (Gupta, 1989). Various statistical methods deal with models in which the observed values are assumed to be independent. However, a quite set of data such as in engineering, natural sciences etc occur in the form of time series where observed values are dependent. However, time series analysis is a systematic approach available to tackle the situation of dependency by observations. Serial correlation coefficient or the autocorrelation coefficient play important role in the analysis of time series which indicates the dependence in its successive values.

Correlogram is a kind of graph that displays the autocorrelation coefficient against the lag period. The process is purely random if the correlogram shows nearly or a zero values for the lag periods, while a value close to one indicates deterministic process (Gupta, 1989). The purpose of analysis of electric consumption is not only to analyze a time series but to generate the data based on the series. Statistical properties are very vital in order to reproduce series of similar statistical characteristics (Gupta, 1989).Electric consumption has a random component being denoted by the word “stochastic”. Stochastic modeling of electric consumption has been widely used for planning and management of power resources and forecasting the occurrence of future consumption. Stochastic models can be used for

forecasting in days, weeks, months and years (Fortin *et al.,* 2004). The previous information can be used as model inputs for forecasting the future event.

Over the last thirty years, time series analysis has become one of the most important and widely used branches of mathematical statistics. Its area of application includes economic forecasting, study of biological data, control systems, signal processing and communications, vibration engineering to mention but few. Unlike multivariate, a single variable time series is referred to as univariate time series. The set of observations that are measured at every instance of time is described as continuous time series. However, time series is taken as discrete when it contains observations being measured at discrete points of time. A series of observations ordered by time is termed as a time series (Granger and Newbold, 1986). Time series are used in many branches of science, technology and businesses. Time series is a function of past values of itself which could be stationary or non-stationary. Unlike non-stationary time series whose economic forecasting is the group of exponentially weighted moving average, stationary time series has constant mean and variance over time.

The smooth non-stationary series can be transformed into a stationary series using a differencing or smoothing filter and then modeled using ARIMA process due to Box and Jenkins (1976). Time series analysis has now become a subject that can be applied in various time based phenomena. Many times, analyses based on the order of movement of variables require the use of equations some of which could be simple or multiple, depending on the nature of the vital features of the variable under consideration (Box and Jenkins, 1976). Weighted methods unlike the ARIMA do not generally need any kind of inferential statistics approach before it can be applied. The two categories of weighted methods; the moving average and exponential smoothing methods which are widely use in commerce and industry have the ability to remove trend and seasonal effects in time series data (Field, 2001). A time

series reflects the dynamic step of the movements ofa phenomenon over a period of time e.g. the figures of price of commodities, national income over a period of time etc.

Any sequence of observations taken on a response (variable) over a period of time is called a time series which can be written, mathematically, as $[t, u(t)]$, *where* $t = 1, 2, 3, ...$ Time series analysis is carried out primarily for the purpose of evaluating past occurrences. Analysis of time series is very important as it helps in studying any phenomenon whose values are or can be arranged in chronological order over a period of time, in making comparative studies in forecasting and it gives approximate indicators. Analysis of time series include; identifying or determining the various factors or influences whose interactions resulted in the variations in the time series; isolating, studying, analyzing and measuring them independently by keeping other factors constant (Gupta, 2004). Time series analysis is very vital not only to businessmen or to economists but also to people who are working in various disciplines such as natural, physical, and social sciences. Time series have about four components among which are; secular trend, seasonal variation, cyclic variation and irregular variation.

Time series analysis have many uses, among which are; it enables the study of past behavior of the phenomenon under consideration; it helps to study the components which are important in the planning of future operations and in the formulation of policy decisions; it helps in carrying out comparative analysis of actual current performance with the expected ones as well as the causes of such variations; it helps to assess the changes in the values of different phenomena at different places. Predictions of future events are called forecasts (Bowerman and O'Connell,1993) and the act of making such predictions is calledforecasting. In many industries predictions of future events known as forecasting is very vital as it paves way for decision-making. In forecastingfuture occurrences, information concerning events that have already occurred inthe past must be realized. In order to have forecasts in place, past information needs to be analyzed in order to identify a pattern that can capture its

important features. Then, this pattern is then extended into thefuture. This forecasting methodis based on the assumption that the pattern being identified will continue in the future to give good forecast, otherwise, it will likely to yield inaccurate predictions (Bowerman and O'Connell, 1993).

It is commonly assumed that the model is the process that generates the series. Nevertheless the yielding model should not be reliable for good forecast; *"All models are wrong but some are useful."*(Box and Jenkins, 1976). Linear models are as less attractive to a forecaster as there appears not a normal distribution or nonlinearity in the series, but in many situations even if there are some elements of nonlinearity in the data the linear model is preferred, this is because the observed non-linearities are not constant over the time or statistically insignificant to improve the linear model-based forecast

3.2 Method of Data Collection

This research work is concerned with fitting time series model to electricity consumption data. The efficient usage of energy is becoming an important issue. Efficiency could be achieved by identifying a reliable solution to the problems of electricity consumption by the fact that electricity consumption forecast is done at a faster rate. The requirement for efficient use of electricity is proper allocation of available resources. This is made possible by accurate prediction of future needs, through its proper estimation process.The data used in this research work was secondary type and collected from Kumbotso Transmission plant, Kano State, covering a period of five years, from 2011 to 2015 and on monthly basis.

3.3 Method of Data Analysis

Analysis of time-series investigates sequences of observations that have been measured at regularly recurring time points. It aims at finding dynamic regularities that may allow for forecasting future occurrences or even controlling the variable (Box & Jenkins, 1974). Unlike short-term forecasting, most linear models such as autoregressive moving average (ARMA)

has less accuracy for long-term forecasting, and usually, it will tend to become uniform. The model is purposely to determine good statistical relationshipsbetween the predicted and historical values of the variables concerned for forecasting (Hendranata, 2003).In this research, and based on certain assumptions; outlier, missing observation, normal distribution and stationary, ARMA and Exponential smoothing methods are used to fit the data. E-viewssoftware is used for the analysis.

3.4 Statistical Parameters

The measures that are computed in order to describe a characteristic of a whole population such as the mean, variance etc of all the observations under consideration are referred to as Parameter.

3.4.1 The Mean

This is the average value of the series, obtained by adding up the series and dividing by the number of observations

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{3.1}$$

where, x_i is the i th number of observations, n is the total number of observations and $\bar{X}$ is the mean of the observations

3.4.2 The Variance

This is a measure of dispersion or spread in the series

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} \tag{3.2}$$

where, n is the number of observations and $\bar{x}$ is the mean of the series

3.4.3 Skewness

This is the measure of asymmetry of distribution of the series around its mean

$$S = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \bar{x}}{\hat{\sigma}}\right)^3 \tag{3.3}$$

where, $\hat{\sigma} = s\sqrt{(n-1)/n}$

3.4.4 Kurtosis

This is the measure of the peakedness or flatness of the distribution of the series

$$K = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \bar{x}}{\hat{\sigma}}\right)^4 \tag{3.4}$$

where, $\hat{\sigma} = s\sqrt{(n-1)/n}$

3.4.5 Serial Correlation Coefficient

This is the measure of the extent to which an observation at a particular time is affected by that at another time.

$$\gamma_j = \frac{n\sum_{i=1}^{n-j}(y_i - \bar{y})(y_{i+j} - \bar{y})}{(n-j)s^2} \tag{3.5}$$

3.5 Jarque-Bera

Jarque-Bera test is a test statistic for testing whether the series is normally distributed. The test statistic measures the difference of the skewness and kurtosis of the series with those from the normal distribution. The statistic is given by

$$JB = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4}\right) \tag{3.6}$$

where, S is the skewness, and K is the kurtosis. Under the null hypothesis of a normal distribution the JB statistic is chi-square distributed with 2 degrees of freedom. The reported probability is the probability that a JB statistic exceeds (in absolute value) the observed value

under the null hypothesis a small probability value lead to the rejection of the null hypothesis of a normal distribution

3.6 Identification of Distribution

The distributions that are normally used in data generation are normal, lognormaletc. The normal distribution is extensivelyused in statistical applications because the sum of variables obtained from anydistribution tends to be distributed normal. To test for normality, the observations are plotted using quantile quantile (Q-Q) plot and if the data points lie on or closer to the diagonal line then the data follows normal distribution. On the other hand, the distribution is normal when its value of coefficient of skewness is close to zero as normal distribution has no coefficient of skewness. However, the distribution follows Log-normal distribution when it is positively skewed, otherwise gamma distribution when moderately skewed.

3.7 Autocorrelation Function (ACF)

It is popularly known that the mean and variance moments are the key properties of random variable. The relationships between observations in different time-periods play a very vital role in time series analysis. These relationships across time periods can be explained by the time series correlation.

3.7.1 Expectation Functions

For a stochastic process, y_t the mean function is defined by

$$\mu_t = E(y_t) \; t = 0, 1, 2,\ldots (3.7)$$

that is, μ_t is the expected value of y_t at time t, where $E(\cdot)$ is the expectation function. μ_t may be different at different time t.

3.7.2 Autocovariance Function

The linear dependence between random variables or observations can be calculated using the autocovariance function. It is defined as

$$\gamma_{t,t+j} = \text{cov}(y_t, y_{t+j})\ t, j = 0,1,2,... \tag{3.8}$$

where,

$$Cov(y_t, y_{t+j}) = E[(y_t - \mu_t)(y_{t+j} - \mu_{t+j})] = E(y_t, y_{t+j}) - E(y_t)E(y_{t+j}) \tag{3.9}$$

The autocorrelation function also describes the linear dependence between observations (as the autocovariance function). The difference is that the autocorrelation function is unitless and bounded, making it somewhat easier to interpret (Cryer, 1986). The autocorrelation function is defined as

$$\rho_{t,t+j} = cor(y_t, y_{t+j})\ t, j = 0,1,2,... \tag{3.10}$$

where,

$$cor(y_t, y_{t+j}) = \frac{\text{cov}(y_t, y_{t+j})}{[\text{var}(y_t)\,\text{var}(y_{t+j})]^{\frac{1}{2}}} \tag{3.11}$$

In order to identify the underlying process, it is useful to check whether these coefficients are statistically nonzero (near zero) or, more specifically, to check whether y_t is a white noise (randomness). Autocorrelation can also be used to identify the order of an AR and a MA process. Under the assumption that $y_t \sim WN(0, \sigma^2)$, the distribution of these coefficients in large samples can be approximated by:

$$\sqrt{T}\rho_{t,t+j} \sim N(1,0) \tag{3.12}$$

Then a usual test of individual significance can be applied

$H_o : \rho_j = 0$
$H_o : \rho_j \neq 0$ For j=1, 2,…

The null hypothesis H_o would be rejected at the 5% level of significance if

$$\left|\rho_{t,t+j}\right| > 1.96\sqrt{T} \tag{3.13}$$

The correlogram plots are usually joined with two-standard error bands around zero, approximated by $\pm 2/\sqrt{T}$. In order to test whether a set of L autocorrelations are jointly zero or otherwise, we test the hypothesis. $H_o : \rho_1 = \rho_2 = ... = \rho_l = 0$.

The most usual test statistic is the Ljung-Box statistic:

$$Q_{LB} = T(T+2)\sum_{J=1}^{M} \frac{\rho_{t,t+j}^2}{T-J} \tag{3.14}$$

Q_{LB} follows asymptotically a chi-square distribution with M degrees of freedom distribution.

3.8 Partial Autocorrelation Function (PACF)

As an important feature of time series, the observations are time dependent and a result of such is the autocorrelation which is the underlying pattern between observations from one time period to the next (Ha *et al*., 2011).

The partial autocorrelation at lag k is defined as the partial correlation between observations k time periods apart. That is, it is the direct correlation between y_t and y_{t-k} after taking out the effect of the intervening observations $y_{t-1}, y_{t-2}, ..., y_{t-k+1}$. It can be estimated using a fast computational algorithm. The collection of autocorrelations at lags $k = 1, 2, 3...$ is known as the autocorrelation function (*ACF*). Similarly, the collection of partial autocorrelations is known as the partial autocorrelation function (*PACF*).The partial autocorrelation plot is commonly used for identification of model (Box and Jenkins).

3.9 Stationarity

It is clear that the estimation of mean, the variance, and the autocorrelation from the *N* data points would involve a problem, since there is need to estimate *N* expected values, *N* variance and covariances. In order to have a solution to this problem, restriction should be formulated for the stochastic process with the assumption that when building a model these

restrictions are satisfied by the process. A stochastic process is said to be stationary if its statistical constants such as the mean and variance remain unchanged.

3.9.1 Moments of Stationary Processes

ARMA process is characterized in terms of the moments of the distribution (mean, ACF and PACF), it is necessary to estimate such using the available data in order to make inference about the underlying process. A consistent estimator of the mean of the process, μ is the sample mean of the time series

$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t \tag{3.16}$$

whose asymptotic variance can be approximated by $\frac{1}{T}$ times the sample variance S_y^2. To check whether the mean of a process is zero or not, the following test is conducted.

$H_o: \ \mu = 0$

$H_a: \ \mu \neq 0$

Under H_o the statistic $\sqrt{T}\,\frac{\bar{y}}{S_y}$ follows approximately a normal distribution.

The concept of stationarity of a stochastic process can be understood as a form of statistical equilibrium (Hipel, 1994). The statistical constants such as mean and variance of a stationary process do not change over time. It is therefore, a necessary condition for building a time series model. Furthermore, the mathematical complexity of the fitted model reduces with this assumption. A time series process $\{x(t), t = 0,1,2,...\}$ is said to be *strongly stationary* or *strictly stationary* if the joint probability distributions function of $\{x_{t-s}, x_{t-s+1}, ..., x_{t+s-1} x_{t+s}\}$ is independent of *t* for all *s*. Therefore for a strong stationary time series process the joint distribution of any set of random variables is time independent (John, 1997).

A stochastic process is said to be *weaklyStationary* of order *k* if the statistical moments of the process up to that order depend upon time differences only and not on the time of occurrences of the data (John,1997). For example a stochastic process $\{x(t), t = 0,1,2,...\}$ is second order

stationary (Hipel, 1994) if it has time independent mean and variance and the covariance values $Cov(x_t, x_{t-s})$ depend only on *s*. It is important to note that neither strong nor weak stationarity implies the other. However, a weekly stationary process following normal distribution is also strongly stationary (John,1997). Some mathematical tests like the one given by Dickey and Fuller (John,1997) are generally used to detect stationarity in a time series data. As mentioned in (Box, 1970) the concept of stationarity is a mathematical idea constructed to simplify the theoretical and practical development of stochastic processes.

In order to design a proper model, efficient for future forecasting, the underlying time series is expected to be stationary. This is not always the case. According to Hipel and McLeod (Hipel, 1994), the greater the time span of historical observations, the greater is the chance for the time series to exhibit stationary characteristics. However, stationary stochastic process can be used to model a time series relatively on short time span. Time series, indicating trend or seasonal patterns are non-stationary in nature. In such a situation, differencing and power transformations are many times used to remove the trend and to make the series stationary.

i) Differencing: A new series is obtained thereby taking difference of order (d). Consequently the new difference data contains less data points than the original data set depending on the required number of differencing. In many cases, first difference is sufficient enough for a series to become stationary.

ii) Line Fitting: A line can be fitted to the data, if there is an indication of trend in the series and then model the residuals obtained there from. The purpose of making a series stationary is to remove its long-term trend and this can be done by fitting a straight line to the data.

iii) Logarithm: This is normally used when there is no constant variance- thereby stabilizing the variance.

3.10 Unit Root Test

Dickey Fuller (DF)test is a kind of formal approach of testing non-stationary in time series data. It is used to test for the presence of unit root in the coefficient of lagged variables. If a value of one is shown by the coefficient of a lagged variable, then the equation indicate that there exists unit root in the series. To test for the presence of a unit root in the data, the following hypotheses are formed.

H_0 : Unit root exist (non- stationary series)

H_a : Unit root does not exist (stationary series)

However, there may be in Dickey Fuller (DF) test a problem of autocorrelation (it assumed that the error term e_t was uncorrelated). To tackle autocorrelation problem, that is, when the error term e_t is correlated, Dickey Fuller have developed a test called Augmented DickeyFuller (ADF) test. This test is conducted by "Augmenting" the equations thereby adding lagged values of the dependent variable (y_t). The number of lagged difference terms to be included is many times obtained empirically, being the idea to include enough terms so as to have the error term in the equations uncorrelated, so that an unbiased estimate of w, the coefficient of lagged variable y_{t-1} can be obtained. For a decision to be taken under ADF, the following equations have to be considered

$$y_t = \beta_0 + wy_{t-1} + e_t \quad (3.17)$$

$$y_t = \beta_0 + \beta_1 t + wy_{t-1} + e_t \quad (3.18)$$

$$y_t = wy_{t-1} + e_t \quad (3.19)$$

Equation (3.17) contains only an Intercept, equation (3.18) contains both Trend and an Intercept while equation (3.19) contains neither Trend nor an Intercept

Decision rule

i. If the absolute value of test-statistics is greater than the absolute value of critical region, we therefore reject null hypothesis

ii. If the absolute value of test-statistics is less than the absolute value of critical region, we therefore accept null hypothesis

iii. If based on the hypothesis test the null hypothesis is rejected, then the data is stationary otherwise non-stationary

3.11 Seasonality or Periodic Fluctuation

Seasonality is a phenomenon that many times occurs when economic time-series like retail sales data are used. If such a pattern is discovered then it must be included in the time-series model. There are different techniques for detecting seasonality, some of which are presented below

i) **The Time Plot:** this displays the data points on graph

ii) **Seasonal Sub-Series Plot:** this is done by plotting the data monthly wise in chronological order then a reference line is drawn at the group means. The autocorrelation plot needs to be checked for periodic fluctuation.

iii) **The Box Plot:** this is also use to detect seasonality in a data by plotting the observed variable on the y-axis against the season on the x-axis in form of a box that is grouping the respective intervals across the period. The 25% lower, median and 75% upper quartile have to be calculated. A box has to be drawn between the quartiles which connect both the minimum and maximum points of the data by a straight line.

iv) **The Autocorrelation Plot:** Seasonality can also be checked by using autocorrelation plot- if a data set is random the autocorrelations should be near zero for all time lags.

Unlike the box plot which shows only the seasonal difference between groups, the sub-series plot shows both the seasonal difference between groups (i.e. for monthly data the difference between January, February, March etc) and that within groups.

3.12 Univariate Time Series Processes

Unlike multivariate, a single variable time series is referred to as univariate time series. AR, MA and ARMA processes are the most common tools used in the study of univariate series as they incorporate the basic ideas of time series analysis.

3.12.1 Stochastic Process

Future values may be predicted from past values when successive observations are statistically dependent. If future values can be determined exactly from past values, it is known as a deterministic process. If, however, predicted values only partly depend on past values, as is the case with most time series data, it is known as a stochastic process, and exact predictions must be replaced by the idea that future values have a probability distribution conditioned on past values.

3.12.2 Autoregressive (AR) Process

An autoregressive model of order *p*, denoted $AR(p)$, satisfies the following equation:

$$y_t = \alpha_1 y_{t-1} + ... + \alpha_p y_{t-p} + e_t \tag{3.20}$$

where, y_t, for $t = 1,2,3,...,T$ are observation values at time t. $\alpha = (\alpha_1, \alpha_2, ..., \alpha_p)$ are constant coefficientsand e_t is the (white) noise term at time $t, e \sim N(0,1)$. As can be seen, the AR process regresses upon itself (on values of its own past values, p, of y_t), hence the prefix `auto' in equation (3.20) can also be expressed using the backshift operator, B

$$y_t - \alpha_1 y_{t-1} - ... - \alpha_p y_{t-p} = e_t \tag{3.21}$$

$$\alpha(B) y_t = e_t \tag{3.22}$$

With $B^p y_t = y_{t-p}$; $p = 1,2,..., p$

where,

$$\alpha(B) = 1 - \alpha_1 B - \alpha_2 B^2 - ... - \alpha_p B^p \tag{3.23}$$

B is the backshift operator on time *t*. For an $AR(p)$ process to be stationary, the roots of $\phi(B) = 0$ must lie within the unit circle (Hamilton, 1994). As an advantage, lag operator does not lie in real notation of the AR and MA models.

3.12.2.1 Stationarity of an AR Process

An AR process is said to be stationary if the roots of the polynomials lie outside the unit circle. This means that if some roots are smaller than unity (ie $|\phi| < 1$), a stationarity may be achieved but if the roots are equals unity then no stationarity is obtainable. A stationary solution is given by the equation below

$$y_t = \sum_{j=0}^{\infty} \phi_j e_{t-j} \tag{3.24}$$

where ϕ_j converges to zero, such that $\sum_{j=0}^{\infty} |\phi_j| < \infty$

3.12.2.2 The ACF of an AR Process

The stationarity of an AR process of high order is given by a mixture of exponential decreasing and sinusoidal components. An exponential decrease in the autocorrelation function is an indication of presence of autoregressive process.

3.12.2.3 The PACF of an AR Process

Once the ACF indicates that an AR model is presence then the PACF is used to determine the order of this AR process. This is because the PACF of AR process tails off to zero at lag greater than or equals $p+1$ (that is to check for any evidence of a departure from zero). This simply means, the PACF "cuts off'' at p

3.12.3 Moving Average (MA) Process

A Moving average model of order *q*, denoted MA (*q*) is generated by a weighted average of current and past shocks. Moving average is a linear regression of the current value of the series against the random shocks of one or more previous values of the series.Fitting a moving average depends on the error terms that are not observable which makes it somehow difficult to fit than an autoregressive model. Parameters estimation in MA model has a less clear interpretation than in the case of AR models and iterative non-linear procedure for model fitting is suggested.

$$y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - ... - \theta_q e_{t-q} \quad (3.25)$$

where, y_t for $t = 1, 2, 3, ..., T$ are the observations at time t and depends on a weighted linear sum of past, q, random shocks e_t (the white noise). These random shocks at particular period are independent of random shocks of other periods while $\theta = (\theta_1, \theta_2, ..., \theta_q)$ are coefficients. Equation (3.25) can be expressed using backshift operator on time *t*

$$y_t = \theta(B) e_t \quad (3.26)$$

where,

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q \quad (3.27)$$

If the roots of $\theta_q(B) = 0$ lie within the unit circle, then the $MA(q)$ process is said to be invertible (i.e y_t can be re-expressed as an infinite order AR process in terms of θ_j for $j = 1, 2, 3, ..., q$. The concept of invertibility is exclusively related to the family of MA models and arises due to the non-uniqueness of the MA autocorrelation functions ($ac.f$) when θ is replaced by θ^{-1} (Cryer, 1986) and can best be described by a simple example using the first order MA model. Consider the following $MA(1)$ models

$$y_t = e_t - \theta e_{t-1} \quad (3.28)$$

$$y_t = e_t - \theta^{-1} e_{t-1} \quad (3.29)$$

It can be shown that these models produce the same ac.f (Cryer 1986) thus it is difficult to uniquely identify a Moving Average (MA) process from a given ac.f. By successive substitution, equations (3.28) and (3.29) can be written in terms of $y_t, y_{t-1}...$ as follows

$$e_t = y_t + \theta y_{t-1} + \theta^2 y_{t-2} + ... \quad (3.30)$$

$$e_t = y_t + \theta^{-1} y_{t-1} + \theta^{-2} y_{t-2} + ... \quad (3.31)$$

If $|\theta| < 1$, the series of coefficients of $y_{t-j}, j > 0$ in equation (3.30) converges whereas in equation (3.31) do not.

$$\theta(B) = 1 - \theta_1(B) - ... - \theta_q(B) = 0 \quad (3.32)$$

A general $MA(q)$ model is invertible if the roots of the equation (3.32) lie within the unit circle (Hamilton 1994, Jones 1987). This is analogous to the stationary restrictions of the coefficients in AR models. Finite-order MA models are always stationary. This is because a MA process resulted from the filtration of a white noise process which exhibit stationarity. Therefore, for a unique MA (q) model to be identified from an ACF, restriction needs to be imposed.

$$\theta = (\theta_1, \theta_2, ..., \theta_q) \quad (3.33)$$

There is no stationarity restriction on the coefficients, but for the process to be invertible. According to this condition of invertibility a MA (q) process is refers to as invertible (Box & Jenkins). Higher order MA models can be defined in similar ways but do however, become complex quite quickly.

3.12.3.1 The ACF of a MA Process

Unlike AR model the ACF of a MA (q) tends to zero for lag $q+1$ or greater. This simply means the ACF "breaks off" at lag q. Therefore, the ACF should be examined and see where it becomes zero for a process to be MA of order q. This can be achieved by using a 95% confidence interval.

3.12.3.2 The PACF of a MA Process

The PACF of a moving average process converges to zero geometrically as such it is helpless in determining the order of MA process.

3.13 Autoregressive Moving Average (ARMA) Models

There are certain assumptions that need to be satisfied before performing the ARMA modeling (Hasmida, 2009); missing observation(s), outlier, stationary and normally distributed series.

3.13.1 Missing Data

This is a situation whereby a missing data point is observed in a data set. The missing data should be replaced by a theoretical algorithm (Yafee and McGee, 2000). Such replacement method is to plug in the mean for the entire series thereby taking the mean of adjacent observations or the mean of the period within which the observation is missing in the series should be used. In some other cases, missing data can be estimated by means of interpolation using linear spines, cubic spines etc

3.13.2 Outlier

Any observation that lies outside the pattern of a distribution is termed outlier (Moore and McCabe, 1999). The presence of an outlier in a distribution is an indication of a sort of problem. Error in measurement or unfitness for the model is a possible cause of outlier. The histograms are easily used to fish out outliers. Any data point (an outlier) should be removed which is a sign of nonstationary series (Hasmida, 2009).

3.13.3 Stationarity

It is clear that the estimation of mean, the variance, and the autocorrelation from the N data points would involve a problem, since there is need to estimate N expected values, N variance and covariances. In order to have a solution to this problem, restriction should be formulated for the stochastic process with the assumption that when building a model these

restrictions are satisfied by the process. A series is needs to be stationary before it can be modeled by using a class of ARMA.

3.13.4 Normal Distribution

A data which follow normal distribution have bell shaped curve with properties such that the curve is symmetric, concentrated in the middle and diminishes on either side. That means there is less chance of extreme values being produced by the data on either side. Non normal data need to be transformed in to normality state by applying log transformation method or Box-Cox transformation method (Hasmida, 2009).

ARMA processes where originally developed by Box and Jenkins 1970 and were investigated by Yule. ARMA models consist of two components; weighted sum of past values (AR) together with the weighted sum of past errors (MA). The ARMA process is stationary and invertible if its roots lie within a unit circle. This condition is in line with conditions imposed on AR (p) and MA (q) processes as in invertible and stationary respectively. An autoregressive moving average model refers to a model with p *AR* terms and q *MA* terms, denoted by $ARMA(p,q)$ and satisfies the following equation

$$y_t = \alpha_1 y_{t-1} + ... + \alpha_p y_{t-p} - \theta_1 e_{t-1} - ... - \theta_q e_{t-q} + e_t \tag{3.34}$$

where, α_p and θ_q are arbitrary constants with $p,q = (1,2,...)$ and $e \sim N(0,1)$ is the white noise error term. For an $ARMA(p,q)$ process to be useful, it must satisfy both the stationary and invertibility constraints.

The autocovariance function of $ARMA(p,q)$

$$\alpha(B) y_t = \theta(B) e_t \tag{3.35}$$

$$y_t = \alpha_1 y_{t-1} + ... + \alpha_p y_{t-p} + \theta_1 e_{t-1} + ... + \theta_q e_{t-q} + e_t \tag{3.36}$$

From equation (3.36)

$$y_t y_{t-k} = \alpha_1 y_{t-1} y_{t-k} + ... + \alpha_p y_{t-p} y_{t-k} + \theta_1 e_{t-1} y_{t-k} + ... + \theta_q e_{t-q} y_{t-k} + e_t y_{t-k} \tag{3.37}$$

By taking expectation, we have

$$\gamma_k = \alpha_1\gamma_{k-1} + ... + \alpha_p\gamma_{k-p} + E(e_t y_{t-k}) + \theta_1 E(e_{t-1}y_{t-k}) + ... + \theta_q E(e_{t-q}y_{t-k}) \quad (3.38)$$

Since $E(e_{t-i}y_{t-k}) = 0, \quad k > i$

$$\gamma_k = \alpha_1\gamma_{k-1} + ... + \alpha_p\gamma_{k-p}, \quad k \geq (q+1) \quad (3.39)$$

And the autocorrelation function (ACF) is given by

$$\rho_k = \alpha_1\rho_{k-1} + ... + \alpha_p\rho_{k-p}, \quad k \geq (q+1) \quad (3.40)$$

By regression theory, we consider predicting y_t based on y_{t-1} and $y_{t-\mu+1}$ as backward can be predicted in time with the same set of predictors, however, the best predictors of $y_{t-\mu}$ is given by

$$\hat{y}_{t-\mu} = \beta_1 y_{t-\mu+1} + \beta_2 y_{t-\mu+2} + ... + \beta_{\mu-1} y_{t-1} \quad (3.41)$$

Similarly y_{t+k} can be predicted from stationary process by regressing on k lagged variables $y_{t+k-1}, y_{t+k-2}, ..., y_t$ that is,

$$y_{t+k} = \phi_{k1} y_{t+k-1} + \phi_{k2} y_{t+k-2} + ... + \phi_{kk} y_t + e_{t-k} \quad (3.42)$$

3.14 The Seasonal ARMA Process

According to Box & Jenkins (1976) and Box *et al.*, (1994), the procedure for fitting seasonal ARIMA (p, d, q) (P, D, Q)[S] process satisfies the following equation

$$\alpha(L)\mathrm{A}(L^s)(1-L)^d(1-L^s)^D y_t = \theta(L)\Theta(L^s)e_t \quad (3.43)$$

when d = D = 0, equation (3.43) becomes ARMA (p, q) (P, Q)[S]

$$\alpha(L)\mathrm{A}(L^s) y_t = \theta(L)\Theta(L^s)e_t \quad (3.44)$$

where P and Q are order of the seasonal autoregressive (AR) and moving average (MA) polynomials while S the seasonal size, is normally taken to be 12 for monthly data and 4 for quarterly data.

$$\mathrm{A}(L^s) = 1 - \mathrm{A}_1 L^s - \mathrm{A}_2 L^{2s} - ... - \mathrm{A}_p L^{ps} \quad (3.45)$$

$$\Theta(L^s) = 1 + \Theta_1 L^s + \Theta_2 L^{2s} + ... + \Theta_q L^{qs} \quad (3.46)$$

The functions A and Θ are the seasonal autoregressive and moving average polynomials

$$\alpha(L) = 1 - \alpha_1 L - \alpha_2 L^2 - ... - \alpha_p L^p \quad (3.47)$$

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q \quad (3.48)$$

The functions α and θ are the autoregressive and moving average polynomials

It assumed that the underlying series is normally distributed, variance is constant and seasonal/non-seasonal relationship is multiplicative (Box *et al.*, 1994). When any or all of these assumptions failed to hold then the series need to be modelled again

Remarks

It is important to note that developing a good model by using both the ACF and PACF plots involve as much as possible trial and error. However, identifying a model with a mixture of both the AR and MA components can be somehow difficult as such the use of information criteria like Akaike information criterion for the identification model is recommended. It is assumed for ARMA model that the series is stationary. A stationary process can be expressed in two parts; deterministic and non-deterministic parts. In a situation whereby a series is seasonal the model can then be extended to include seasonal autoregressive and seasonal moving average terms. A data set consisting of at least 50 observations is recommended for ARMA model (Chatfield, 2001) while other authors recommend at least 100 observations. ARMA model is more parsimonious than pure AR and MA processes (minimizes the number of parameters)

3.15 Akaike's Information Criterion (AIC)

(Akaike, 1973) Akaike Information Criterion (*AIC*) was initially proposed by Akaike. Akaike Information Criterion is a measure of the relative quality of statistical models for a given set of data based on the principle of parsimony. It proposes the use of relative entropy, or the Kullback-Leibler Information (KLI), as a means for model selection. The first term is a

measure of lack of fit, while the second term measures the increase in the order of the model, thereby having parsimony in the number of parameters being considered. The *AIC* Criterion to be minimized is given by

$$AIC = \log(\hat{\sigma}_k^2) + \frac{2k}{T} \tag{3.49}$$

where, k is the dimensionality of the model, $\hat{\sigma}_k^2$ is the maximum likelihood estimate of thewhite noise variance, and T is the sample size. Akaike's Information Criterion *AIC* has been used in many areas such astime series analysis e.g. mixture distribution modeling (Raijmakers, 2001), luminance detectionmodeling (Smith, 1998), and perception modeling using stochastic catastrophe models (Ploeger, 1998).

3.15.1 Corrected Akaike Information Criterion

Corrected *AIC* (*AIC c*), proposed by Hurvich and Tsai, 1989 is a bias-corrected version of *AIC* (Brockwell and Davis, 2002). This criterion takes into account the *AIC* 's deficiency when large number of parameters in relation to the sample size is involved (Sugiura, 1978). The *AIC c*Criterion to be minimized is given by

$$AICc = \log(\hat{\sigma}_m^2) + \frac{2(m+1)}{T-m-2} \tag{3.50}$$

where, $m, \hat{\sigma}_m^2$ and T as defined in (3.49). The penalty term $\frac{2(m+1)}{T-m-2}$ in(3.50) avoid the tendency of data over-fit by *AIC* .

3.16 Model Parsimony

In time series modeling the principle of parsimony has to be considered (Zhang, 2007). This states that the model with smallest number of parameters is to be used (Chatfield, 1996). The simplest among the competing models should be considered, still maintaining an accurate description of inherent properties of the time series. The idea of model parsimony is similar

to the famous *Occam's razor* principle (Hipel, 1994). As discussed by Hipel and McLeod (Hipel, 1994) one aspect of this principle is that when face with a number of competing and adequate explanations, pick the most simple one. The Occam's razor provides considerable inherent informations, when applied to logical analysis.

Moreover, the more complicated the model, the more possibilities will arise for departure from the actual model assumptions. With the increase of model parameters, the risk of overfitting also subsequently increases. An over fitted time series model may describe the training data very well, but it may not be suitable for future forecasting. As potential overfitting affects the ability of a model to forecast well, parsimony is often used as a guiding principle to overcome this issue. Thus in summary it can be said that, while making time series forecasts, genuine attention should be given to select the most parsimonious model among all other possibilities

3.17 The Model Building Process

The process of choosing an appropriate model for a given time series is called the model building process, and can be broken up into three distinct stages (Chatfield, 2004); model identification, model fitting (estimation) and model verification (diagnostic checking).

3.17.1 Model Identification

Identification of model involves examining the given data using various statistical methods. A single statistic of the data may be efficient; however multiple statistics would certainly give a firmer basis for any postulates to be made. Knowledge of the subject area from which the data arose such as economics, biology, or chemistry would also be applied. For example in areas such as oceanography, long stationary series may occur (Chatfield, 2004). The gathered information and computed statistics will ultimately lead one to postulating a class, or a range of classes, of suitable models. For example, examining the autocorrelation or partial autocorrelation function of a data set may lead to choosing the class of AR or MA models as

the most suitable class. Therefore correlogram and partial correlogram aid in choosing appropriate model through finding out the values of *p,d,*and *q.*

3.17.2 Parameter Estimation (Model Fitting)

Once a class (or classes) of models have been chosen, that model's parameters need to be estimated so as to fit the model to the given data. Thus, for example, if we think an *AR* model of order 1 may adequately fit the data, we would require two parameters to be estimated, namely α_1 and σ^2. There are many methods available for estimation of model parameters and their selection depends on the process under investigation. For pure *AR* process, parameter estimation methods include Moments, Least Squares, Ordinary Least Squares (Engle, 1985), Maximum Like-lihood, Approximate Maximum Likelihood, Yule-Walker (Cryer, 1986) the covariance method. For pure *MA* processes, parameter estimation methods include Moments, the innovations algorithm (Brockwell, 2002), and Maximum Likelihood. For *ARMA* processes, one could use Moments, the Long *AR* method, Hannan-Rissanen procedure (Rao, 1993) or Maximum Likelihood.

3.17.3 Diagnostic Checking (Model Verification)

The model verification deals with assessing the residuals of the model in order to decide whether to accept the model, or not. If evidence results in the latter case, whether it is as a result of additional data or inadequacies of the model, the building process of the model will need to be repeated from step 2, or even from the beginning. Repeating of this process may oftenly occur before a decision is finally taken about the model. Therefore model building is an interactive and iterative process (Chatfield, 2004).

Remarks

It must be remembered that the model that is finally decided upon is only an approximation to the "true" model that underlies the data. It could in fact turn out that there is no true model and that our approximation is only an approximation to the observed behaviour of the data,

and nothing more. This, however, does not discredit the importance of modeling the data as accurately as possible.

3.17.3.1 Diagnostic Testing for Parameter

Statistical test of significance of the AR and MA parameters should be carried out to check if the model is overspecified (t-statistic test).On the other hand, both the stationarity and invertibility conditions are needed to be satisfied. Factorizing the AR and MA polynomials yield the following:

$$\alpha(L) = (1-R_1^{-1})(1-R_2^{-1})...(1-R_p^{-1}) \quad (3.51)$$

$$\theta(L) = (1-S_1^{-1})(1-S_2^{-1})...(1-S_q^{-1}) \quad (3.52)$$

Having any of these roots close to one, it may be an indication of non-stationarity and/or vertibility. A close inspection of the covariance matrix of the estimated parameters will pave a way for detecting the possible presence of high correlation between the estimates of some parameters which can be an indication of the presence of a 'common factor' in the model (Box and Jenkins; 1976).

3.17.3.2 Diagnostic Testing for Residuals

Once a candidate model has been adequately identified, then there is need for its residuals to be approximately white noise that is if the residuals have zero mean and are uncorrelated. The key tools for such analysis are the time plot, the ACF and the PACF of the residuals. The theoretical ACF and PACF of white noise processes take value zero for lags $j \neq 0$, so if the model is adequate, then high percentage of the coefficients of the sample ACF and PACF should be close to zero respectively. In practice, there is the need to have about 95% of these coefficients fall within the non-significance bounds. Also, the Ljung-Box statistic (3.9) should assume small values, as corresponds to uncorrelated variables. This test statistic follows chi-square $(m-k)$ with $k = p+q$. In a situation whereby the model is not

appropriate, that is the correlograms (simple and partial) of the residuals depart from white noise,then there is need for the reformulation of the model

3.18 Comparison of Models and Forecast Evaluation

Let $y_1, y_2, y_3, ..., y_n$ be a univariate time series observations and wants to forecast for x_{n+h} , $h = 1, 2, \ldots$ then a data point, $\hat{y}_n(h)$ can be forecasted which depends on past and present data points which possibly augmented with a function of time (Chatfield, 2001). However, forecasting model a mathematical representation of real events defers from forecasting method a procedure for computing a forecast.The model is adequate for forecasting if it is scaled through the entire diagnostic test. This final model is obtainable by having its accuracy higher than any other competing models which confirmed it is performance in term of in-sample (or out-sample) forecast. In this research work the obtained forecast is tested against three standard measures: The mean absolute error (MAE), root mean squared error (RMSE) and loss function. Let $x_1, x_2, ..., x_n$ be actual observations and $\bar{x}_1, \bar{x}_2, ..., \bar{x}_n$ be random variable of x then

$$MAE = \frac{1}{n}\sum_{i=0}^{n}(x_i - \bar{x}_i) \tag{3.53}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(x_i - \bar{x}_i)^2} \tag{3.54}$$

Loss function suggested and defined as a forecast that minimizes average loss when divided over the probability distribution of forecast error. To evaluate a forecast a loss function $L(e)$ can also be defined in association with the forecasting error, such that ($e = observed\ \ value - forecast$) Loss function has two usual properties;

a) $L(0) = 0$

b) $L(e)$ increases with absolute value of e

Quadratic loss function is given by

$$L(e) = \delta_1 e^2 \tag{3.55}$$

Absolute error loss function is given by

$$L(e) = \delta_2 |e| \tag{3.56}$$

where, δ_1 and δ_2 are constants respectively

3.19 Static and Dynamic Forecasting

Unlike the dynamic forecasting in which forecast are obtained thereby using the forecasted value for the periods after the first period (the lagged variable), the Static forecasting uses the actual data point rather than the forecasted value for the lagged variable (EViews)

3.20 Model Based Forecasting: After a particular model has satisfied all the required conditions; identification, parameter estimation and diagnosis there need for use of it to forecast future occurrences from the fitted model. In such a situation a quadratic loss function can be used;

$$\hat{y}_n(h) = E(y_{n+h} \mid Z, I_n) \tag{3.67}$$

where, $\hat{y}_n(h)$ is the forecast value, Z is the model and I_n is the information at time n

3.20.1 Autoregressive (AR) Models

$$y_t = \alpha_1 y_{t-1} + ... + \alpha_p y_{t-p} + e_t \tag{3.58}$$

One-step ahead forecastis given by

$$\hat{y}_{t-1}(1) = \hat{\alpha}_1 y_{t-1} + ... + \hat{\alpha}_p y_{t-p} \tag{3.59}$$

where, the true coefficients are replaced by in-sample estimates $\hat{\alpha}_j$

Two-step ahead forecastis given by

$$\hat{y}_{t-1}(2) = \hat{\alpha}_1 \hat{y}_{t-1}(1) + \hat{\alpha}_2 y_{t-1} + ... + \hat{\alpha}_p y_{t-p+1} \tag{3.60}$$

where the unknown observation y_t is replaced by its prediction $\hat{y}_{t-1}(1)$

3.20.2 Moving Average (MA) Models

Visual inspection of ACF or information criteria can be used to identify the lag order of moving average process. This helps in the estimation of the coefficient, θ_j. Given a Moving average model

$$y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - ... - \theta_q e_{t-q} \tag{3.61}$$

One-step ahead forecastbased on the availability of $\hat{\theta}_j$ and $\hat{e}_t$ is given by

$$\hat{y}_{t-1}(1) = \hat{\theta}_1 \hat{e}_{t-1} - \hat{\theta}_2 \hat{e}_{t-2} - ... - \hat{\theta}_q \hat{e}_{t-q} \tag{3.62}$$

where, the true coefficients are replaced by in-sample estimates $\hat{\theta}_j$

Two-step ahead forecastis given by

$$\hat{y}_{t-1}(2) = \hat{\theta}_2 \hat{e}_{t-1} - \hat{\theta}_3 \hat{e}_{t-2} - ... - \hat{\theta}_q \hat{e}_{t-q+1} \tag{3.63}$$

The forecast will after some steps becomes zero

3.20.3 Autoregressive Moving Average (ARMA) Model

Autocorrelation and partial autocorrelation functions approach cuts off to zero geometrically for ARMA processes. However, Lag orders p and q are difficult to be identified by visual inspection of ACF and PACF respectively. Therefore, information criteria should be used to set the lag orders. ARMA models could be appropriate when especially some parameters are zero (unlike a situation whereby exact models including all parameters are concluded, it could be better of prediction if the parameters differ from zero)

One-step ahead forecastbased on the availability of $\hat{\alpha}_j$, $\hat{\theta}_j$ and $\hat{e}_t$ is given by

$$\hat{y}_{t-1}(1) = \hat{\alpha}_1 y_{t-1} + ... + \hat{\alpha}_p y_{t-p} - \hat{\theta}_1 \hat{e}_{t-1} - \hat{\theta}_2 \hat{e}_{t-2} - ... - \hat{\theta}_q \hat{e}_{t-q} \tag{3.64}$$

Two-step ahead forecastis given by

$$\hat{y}_{t-2}(2) = \hat{\alpha}_1 \hat{y}_t(1) + \hat{\alpha}_2 y_{t-2} + ... + \hat{\alpha}_p y_{t-p} - \hat{\theta}_2 \hat{e}_{t-1} - ... - \hat{\theta}_q \hat{e}_{t-q+1} \tag{3.65}$$

3.21 Exponential Smoothing

Exponential smoothing is an adaptive forecasting method. It is an effective means of making forecast when there are a few observations on which to base the forecast.Forecasts from exponential smoothing unlike forecasts from regression models where fixed coefficients are used, are adjust based upon past forecast errors.

3.21.1 Single Exponential Smoothing

This method is appropriate for series that fluctuate randomly above and below a constant mean with neither a trend nor seasonal pattern. Mathematically,

$$\hat{y}_t = \phi y_t + (1-\phi)\hat{y}_{t-1}, \qquad 0 < \phi < 1 \tag{3.66}$$

The smaller the value of ϕ the smoother is the $\hat{y}_t$. Bowerman and O'Connell (1979) suggest that values of ϕ around 0.01 to 0.30

3.21.2 Double Exponential Smoothing

This is appropriate for series with a linear trend. The single smoothing method is applied twice using the same parameter.

$$S_t = \phi y_t + (1-\phi)S_{t-1} \tag{3.67}$$

$$D_t = \phi S_t + (1-\phi)D_{t-1} \tag{3.68}$$

where, S is the single smoothed and D is the double smoothed series respectively

3.21.3 Holt-Winters-Multiplicative

This method is appropriate in a situation where there is series with a linear time trend and multiplicative seasonal variation

$$\hat{y}_{t+k} = (a + bm)c_{t+m} \tag{3.69}$$

where, a is the intercept, b is the trend and c_t is the multiplicative seasonal effect

$$a(t) = \phi \frac{y_t}{c_t(t-s)} + (1-\phi)(a(t-1) + b(t-1))$$

$$b(t) = \beta(a(t) - a(t-1)) + (1-\beta)b(t-1)$$

$$c_t(t) = \gamma \frac{y_t}{a(t)} + (1-\gamma)c_t(t-s)$$

The forecasts can be computed by using

$$\hat{y}_{t+k} = (a(T) + b(T)k)c_{T+k-s} \tag{3.70}$$

3.21.4 Holt-Winters-Additive

This method is appropriate in a situation where there is series with a linear time trend and additive seasonal variation

$$\hat{y}_{t+k} = a + bk + c_{t+k} \tag{3.71}$$

where, a is the intercept, b is the trend and c_t is the additive seasonal effect

$$a(t) = \phi(y_t - c_t(t-s)) + (1-\phi)(a(t-1) + b(t-1))$$

$$b(t) = \beta(a(t) - a(t-1)) + 1 - \beta b(t-1)$$

$$c_t(t) = \gamma(y_t - a(t+1)) - \gamma c_t(t-s)$$

The forecasts can be computed by using

$$\hat{y}_{T+k} = a(T) + b(T)k + c_{T+k-s} \tag{3.72}$$

CHAPTER FOUR

RESULTS AND DISCUSSION

Data Analysis and Discussion of Findings

4.1 Introduction

In this chapter data is being analyzed, discussed and commented using ARMA and exponential smoothing modeling methods for forecasting of electricity consumption (mwh). The models will be checked to get a suitable model to fit the data. Data points from January 2011 to February 2015 are used as a training set while data points from March 2015 to December 2015 are used for validation.

4.2 Data Analysis

The data set was found to be stationary, distributednormal and free from missing value and outlying observation(s) respectively.

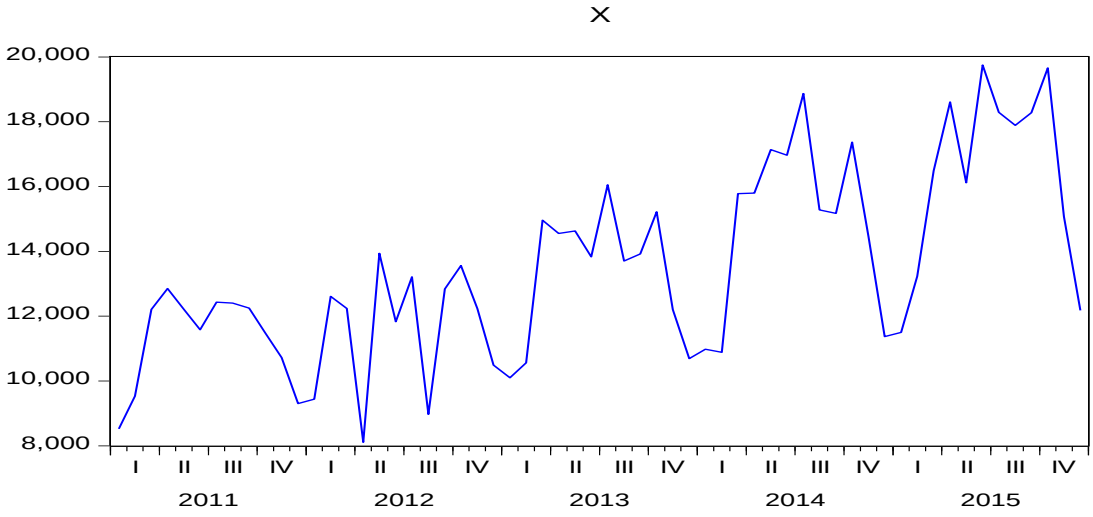


Figure 1: Time plot of the data

The time plot displayed in figure 1 suggests a stationary series. Nevertheless, the unit root test (ADF test) yields an ADF statistics value of minus 5.132 (0.001) corresponding to the critical value of minus 3.500 at 5% level of significance. This leads to the rejection of null hypothesis of unit root exist that the series is stationary as the absolute value of ADF statistics is greater than the absolute value of the critical value at level.

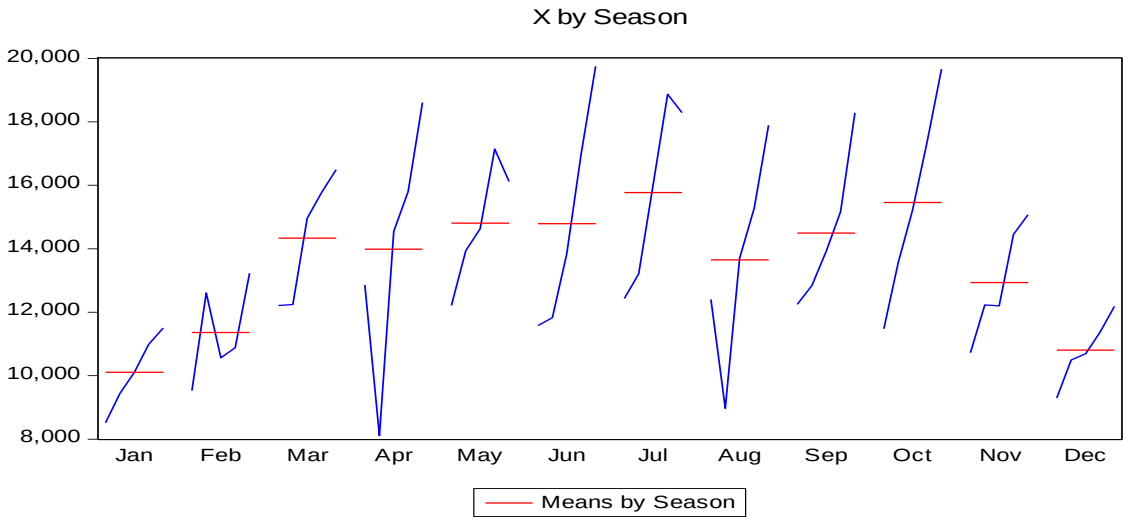


Figure 2: Seasonal effects

Figure 2, shows that the data exhibit seasonality having the minimum value in January while maximum in July.

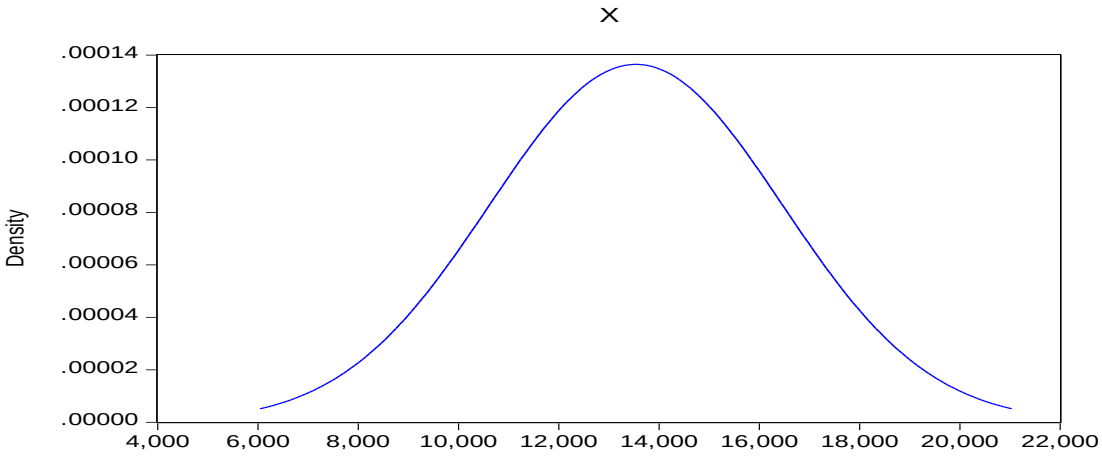


Figure 3: Theoretical distribution plot

The theoretical distribution plot in figure 3 show that the series is approximately normally distributed

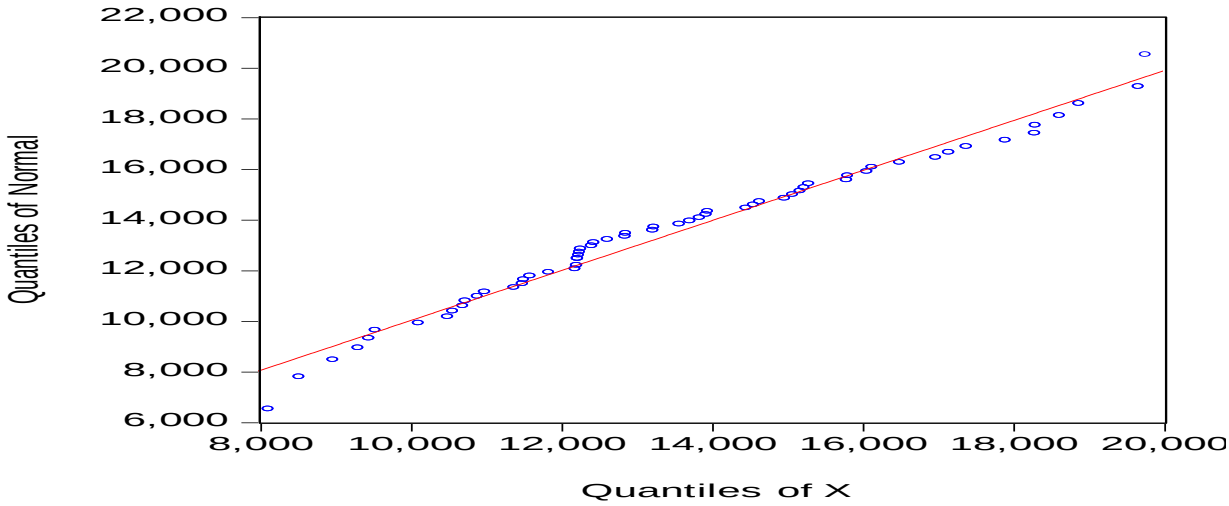


Figure 4: Q-Q plot of the series

The Q-Q plot in figure 4show that the data is normally distributed as points lay approximately closer to the diagonal line

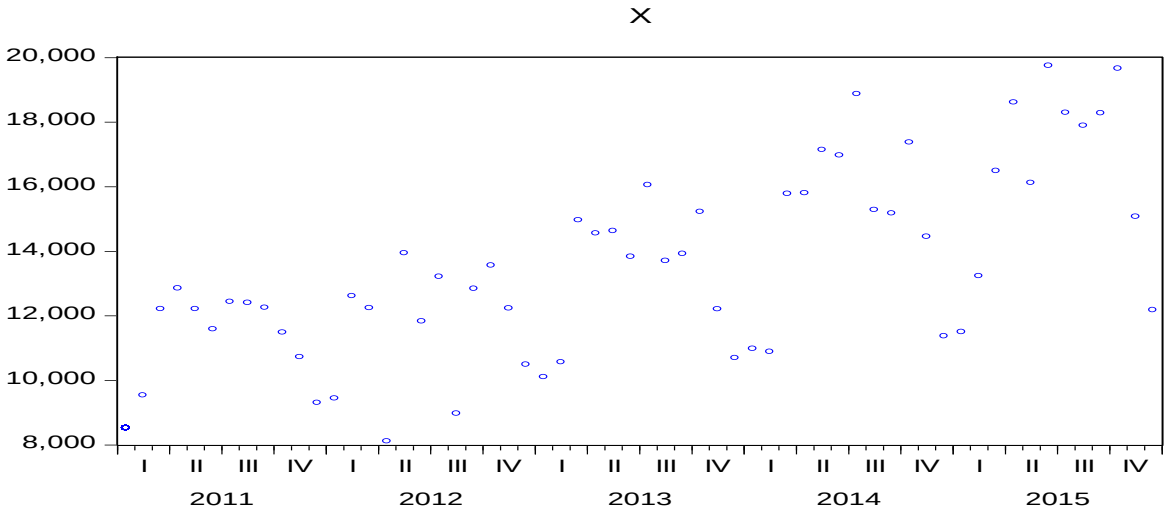


Figure 5: Dot plot of the series

The dot plot in figure 5 indicate that the series contained no outlying observation (s)

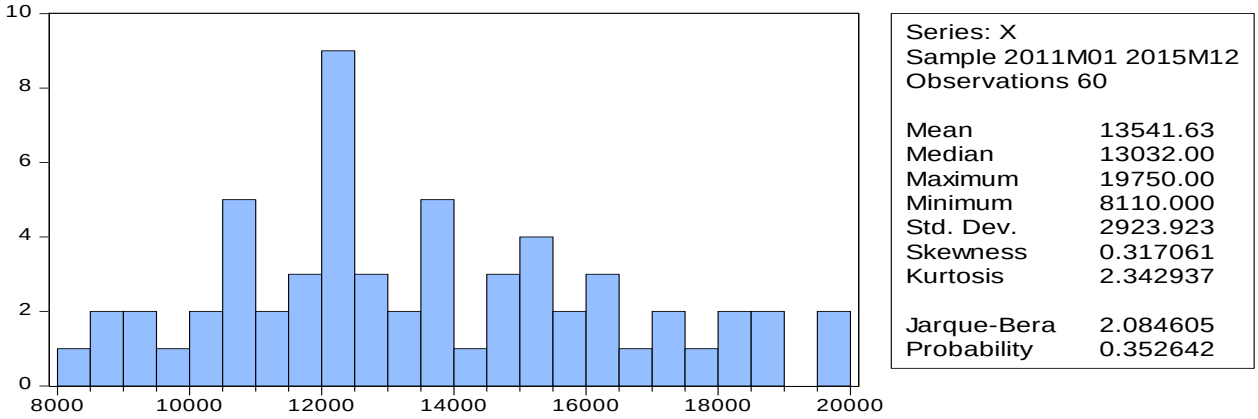


Figure 6: Histogram and descriptive statistics of the series

TheJarque-Bera (JB) test in figure 5 concludes that the null hypothesis of normal distribution would not be rejected at 5% level ($0.3526 > 5\%$)

Table 1: Descriptive statistics of electric consumption (mwh)

Statistic	Mean	Median	Max	Min	Std. Dev.	Skewness	Kurtosis	Sum Sq. Dev.
Value	13541.63	13032.00	19750.00	8110.00	2923.92	0.3171	2.3430	5.04E+08

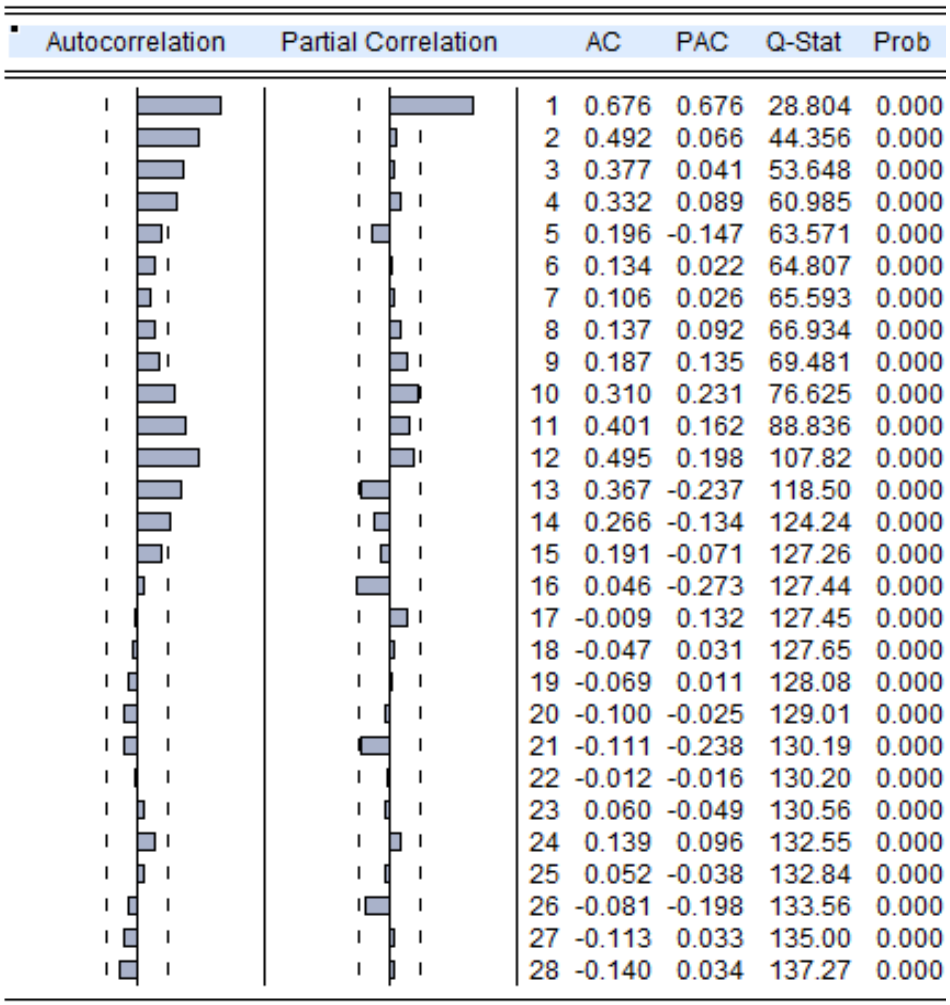
The summary or descriptive statistics of the data is given in table 1 above

4.2.1 Model Identification

Autocorrelation is the basis for model identification (model fitting). The autocorrelation function is visually examined for important features, behavior and pattern of autocorrelation across lags. This enable conclusion about suitable model to be drawn. Furthermore a series is said to be random if all the autocorrelation values set to or close to zero. A series is non-stationary when there is no decay in the autocorrelation to zero.

Table 2: Theoretical pattern of identification of model

S/N	ACF	PACF	Model
1	Rapid decay to zero	Breaks off at lag k	AR(k)
2	Oscillating decay to zero	Breaks off at lag k	AR(k)
3	Rapid/oscillating decay	Breaks off at lag p	AR(p)
4	Breaks off at lag k	Oscillating decay to zero	MA(k)
5	Breaks off at lag k	Rapid decay to zero	MA(k)
6	Breaks off at lag q	Rapid/oscillating decay	MA(q)
7	Decay begins at lag q	Decay begins at lag p	ARMA(p , q)



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.676	0.676	28.804	0.000
		2	0.492	0.066	44.356	0.000
		3	0.377	0.041	53.648	0.000
		4	0.332	0.089	60.985	0.000
		5	0.196	-0.147	63.571	0.000
		6	0.134	0.022	64.807	0.000
		7	0.106	0.026	65.593	0.000
		8	0.137	0.092	66.934	0.000
		9	0.187	0.135	69.481	0.000
		10	0.310	0.231	76.625	0.000
		11	0.401	0.162	88.836	0.000
		12	0.495	0.198	107.82	0.000
		13	0.367	-0.237	118.50	0.000
		14	0.266	-0.134	124.24	0.000
		15	0.191	-0.071	127.26	0.000
		16	0.046	-0.273	127.44	0.000
		17	-0.009	0.132	127.45	0.000
		18	-0.047	0.031	127.65	0.000
		19	-0.069	0.011	128.08	0.000
		20	-0.100	-0.025	129.01	0.000
		21	-0.111	-0.238	130.19	0.000
		22	-0.012	-0.016	130.20	0.000
		23	0.060	-0.049	130.56	0.000
		24	0.139	0.096	132.55	0.000
		25	0.052	-0.038	132.84	0.000
		26	-0.081	-0.198	133.56	0.000
		27	-0.113	0.033	135.00	0.000
		28	-0.140	0.034	137.27	0.000

Figure 7: Correlogram of the data at level

It observed based on the correlograms in figure 7 that the autocorrelations exhibit an oscillating decay while the partial autocorrelation indicates a break off at lag 1. This conclude

and suggests based on theoretical pattern of model identification an AR(1) and hence $(p+q \leq 2)$. In this research work, Akaike's information criterion was employed for optimizing the model selection.

4.2.2 Parameter Estimation

Table 3: Model identification

Model	**RSS**	**AIC**	**Normality test**		**Serial Correlation**		**ARCH Effect**	
			JB Test	**P Value**	**BG Test**	**P Value**	**LM Test**	**P Value**
AR(1)	2.48E+08	18.1556	0.7962	0.6716	0.1821	0.8340	0.0007	0.9790
MA(1)	3.35E+08	18.4405	1.0018	0.6060	7.4921	0.0013	0.4769	0.4926
ARMA(1,1)	2.46E+08	18.1830	0.7576	0.6847	0.0186	0.9816	0.0284	0.8669
SAR(1)	1.52E+08	17.9558	8.8808	0.0118	0.1295	0.8789	0.0861	0.7706
SMA(1)	2.44E+08	18.1567	0.7200	0.6977	5.5069	0.0066	0.3592	0.5513
SARMA(1,1)$(1,1)^{12}$	9.22E+07	17.5397	3.4768	0.1758	3.1008	0.0560	0.0530	0.8190

It is observed from table 3 that SARMA (1,1) (1,1)[12] outperformed the remaining competing models. However, the null hypothesis of white noise cannot be rejected, because the p-value is insignificant at 5% level .The null hypothesis of Heteroskedasticity is rejected, because the p-value is significant at 5% level.

Table 4: Parameters of the proposed model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	39952.26	46594.10	0.857453	0.3961
AR(1)	0.762486	0.087227	8.741395	0.0000
SAR(12)	0.943375	0.100015	9.432290	0.0000
MA(1)	-0.972103	0.016113	-60.32939	0.0000
SMA(12)	-0.838021	0.145314	-5.766981	0.0000
R-squared	0.759594	Akaike info criterion		17.53965
Adjusted R-squared	0.736698	Durbin-Watson stat		1.481194
F-statistic	33.17614	Prob(F-statistic)		0.000000

4.2.3 Diagnostic Checking

Having the least value of RSS and AIC, SARMA (1,1) (1,1)[12] outperformed the remaining competing models. The residuals of the estimated model are assumed to be white noise, homoscedastic and normally distributed. Failure in any of these assumptions leads to the inefficiency of the model but yet consistent unless the residuals are not white noise. However, if residuals are not white noise and heteroscedastic then the standard errors of the estimated parameters are said to be biased and model has to be modified. On the other hand, the t-statistics are not t-distributed if residuals are not normally distributed. The Autocorrelation in figure 9 above are very small as no one exceeds the confidence bound. This indicates the presence of white noise.

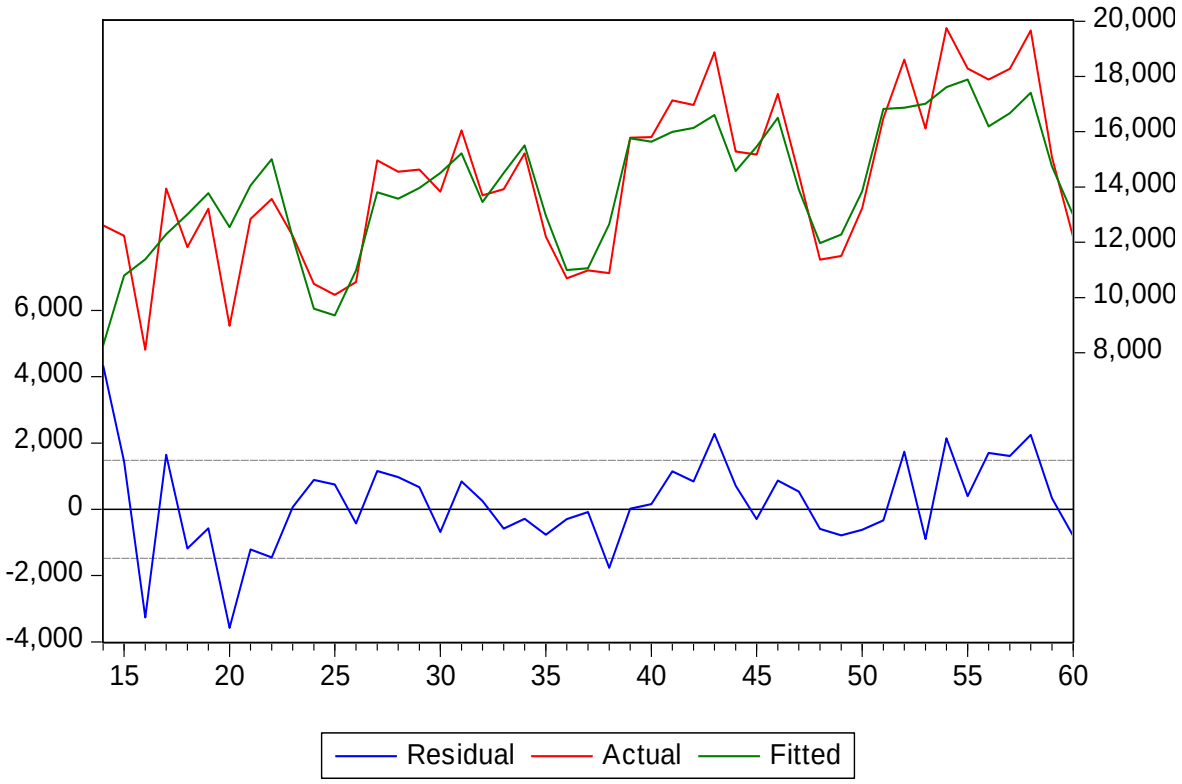
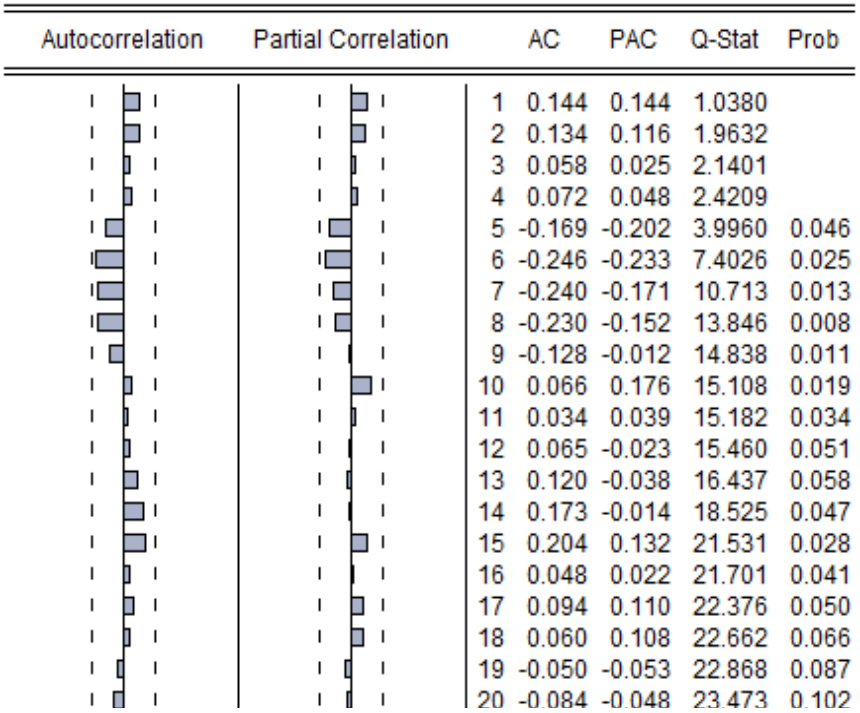


Figure 8: Residuals plot of the proposed model



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.144	0.144	1.0380	
		2	0.134	0.116	1.9632	
		3	0.058	0.025	2.1401	
		4	0.072	0.048	2.4209	
		5	-0.169	-0.202	3.9960	0.046
		6	-0.246	-0.233	7.4026	0.025
		7	-0.240	-0.171	10.713	0.013
		8	-0.230	-0.152	13.846	0.008
		9	-0.128	-0.012	14.838	0.011
		10	0.066	0.176	15.108	0.019
		11	0.034	0.039	15.182	0.034
		12	0.065	-0.023	15.460	0.051
		13	0.120	-0.038	16.437	0.058
		14	0.173	-0.014	18.525	0.047
		15	0.204	0.132	21.531	0.028
		16	0.048	0.022	21.701	0.041
		17	0.094	0.110	22.376	0.050
		18	0.060	0.108	22.662	0.066
		19	-0.050	-0.053	22.868	0.087
		20	-0.084	-0.048	23.473	0.102

Figure 9: Correlogram of residuals of the proposed model

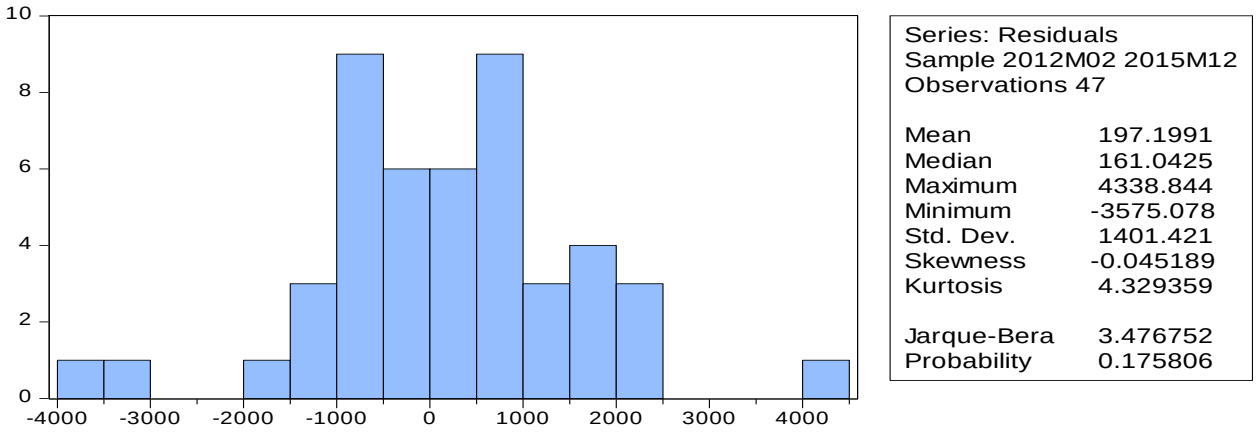


Figure 10: Histogram and descriptive statistics of residuals

According to the Jarque-Bera test in figure 10 above, the null-hypothesis of normality cannot be rejected; the kurtosis being one more than three with an approximate zero value of skewness. This is because the p-value is insignificant at 5% level. More so, the kurtosis is fairly greater than three. In conclusion, the residuals have satisfied all the assumptions mentioned above.

4.3 Exponential Smoothing

Table 5: Exponential smoothing methods

S/N	Method	ϕ	β	γ	RSS	RMSE	Mean	Trend
1	Single exponential	0.6640			2.90E+08	2198.9	13614.2	
2	Double exponential	0.0300			2.95E+08	2217.5	16555.5	94.74
3	Holt-winter Additive	0.0500	0.0000	0.0000	9.55E+07	1261.9	16856.0	106.92
4	Holt-winter Multiplicative	0.0200	0.0000	0.0000	8.54E+07	1192.9	16811.9	106.92
5	Holt-winter No seasonal	0.7200	0.0000		2.85E+08	2181.2	13414.9	251.00

Varieties of exponential smoothing; single exponential, double exponential, Holt-winters additive, Holt-winters multiplicative and Holt-winters no seasonal are compared on electricity consumption (MWH). The result shows that Holt-winters multiplicative at 2% outperformed other members thereby minimizing the sum of square residuals of the data best. This leads to the rejection of the hypothesis that the series is free from trend and seasonal effects

Table 6: Dynamic and Static forecasting

		RMSE	MAE	TIC	VP	CVP
Dynamic	In Sample	1475.3	1056.6	0.05	0.02	0.98
	Out sample	1278.3	1105.4	0.04	0.55	0.40
Static	In Sample	1412.9	1038.4	0.05	0.07	0.93
	Out sample	1278.3	1105.3	0.04	0.55	0.40

It is observed from table 5 that Holt-winters multiplicative can be used for both dynamic and static forecasting

4.4 Model Comparison

Table 7: Forecast evaluation

	Model	**RSS**	**RMSE**
1.	ARMA (1,1)(1,1)[12]	9.22E+07	1278.3
2.	Holt-winter multiplicative	8.54E+07	1192.9

The two competing methods are based on loss function compared in table 7. The results show that Holt-winters multiplicative minimize the sum of square residuals of the series better than seasonal ARMA.

4.5 Contribution to Knowledge

- The research has been able to use Autoregressive Moving Average (ARMA) and Exponential smoothing methods to study and forecast electricity consumption (mwh); A case study of Gazaoua power plant
- The research has been able to check for stationarity of the data by using Augmented Dickey Fuller test and ACF
- The research has successfully compared the forecasting ability based on loss function of Autoregressive Moving Average (ARMA) and Exponential smoothing methods on electricity consumption (mwh)

CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMMENDATION

5.1 Summary

Augmented Dickey fuller test was performed to check for presence of unit root and was found that the series was stationary with an indication of trend and seasonal effect. The study compared the forecasting ability of Autoregressive and exponential smoothing techniques on electricity consumption (mwh). However, loss function was used as a basis for the comparison and hence Holt-winters multiplicative outperformed seasonal autoregressive moving average.

5.2 Conclusion

The objectives of the study is to propose a forecasting model for monthly electricity consumption using ARMA and Exponential Smoothing approaches in order to investigate their forecasting ability. Based on the findings, the tentative competing models that satisfied the criteria are Seasonal ARMA (1,1) (1,1)[12] and Holt-winters multiplicative. However, the results show that Holt-winters multiplicative emerged as the best forecasting method.

5.3 Recommendation

The result show an increase in the consumption which is probably as a result of rapid increase in population therefore, government should make proper arrangement in order to face the challenge thereby providing basic devices.

APPENDIX

RELATIONSHIP BETWEEN AUTOREGRESSIVE (AR) AND MOVING AVERAGE (MA) PROCESSES

- **THE MOVING AVERAGE (MA) PROCESS**

An $AR(1)$ process, $y_t = \alpha_1 y_{t-1} + e_t$ can be expanded to give MA process that is

$y_t = \alpha_1 y_{t-1} + e_t$ (1)

It implies that $y_{t-1} = \alpha_1 y_{t-2} + e_{t-1}$ (2)

Substituting (2) into (1), we have

$$y_t = \alpha_1(\alpha_1 y_{t-2} + e_{t-1}) + e_t$$

$$y_t = \alpha_1^2 y_{t-2} + \alpha_1 e_{t-1} + e_t \quad (3)$$

In general, $y_t = \alpha_1^t y_0 + \alpha_1^{t-1} e_1 + ... + \alpha_1 e_{t-1} + e_t$ (4)

$$y_t = \alpha_1^t e_0 + \alpha_1^{t-1} e_1 + ... + \alpha_1 e_{t-1} + e_t \quad (5)$$

Equation (5) above is $MA(\infty)$ as $t \to \infty$ with $-\theta_s = \alpha_1^s$ for $s = 1, 2, 3, ..., t$

- **THE AUTOREGRESSIVE (AR) PROCESS**

An $MA(1)$ process, $y_t = e_t - \theta_1 e_{t-1}$ can be expanded to give AR

$$y_t = e_t - \theta_1 e_{t-1} \quad (6)$$

$$y_t = e_t - \theta_1(y_{t-1} + \theta_1 e_{t-2})$$

$$y_t = e_t - \theta_1 y_{t-1} - \theta_1^2 e_{t-2}$$

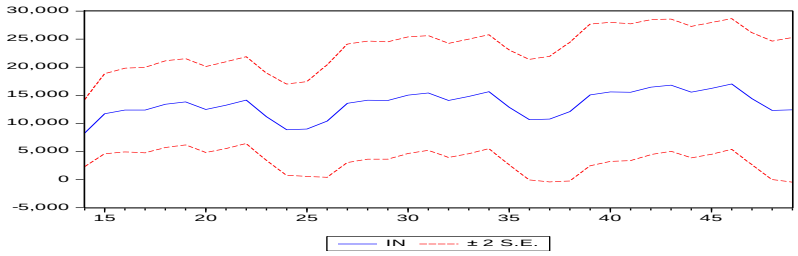
$$y_t = e_t - \theta_1 y_{t-1} - ... - \theta_1^{t-1} y_1 - \theta_1^t e_0 \quad (7)$$

In general, $y_t = e_t - \theta_1 y_{t-1} - ... - \theta_1^{t-1} y_1 - \theta_1^t y_0$ (8)

Equation (8) is equivalent to $AR(\infty)$ process as $t \to \infty$ with $\alpha_s = -\theta_1^s$ for $s = 1, 2, 3, ..., t$

FORECAST

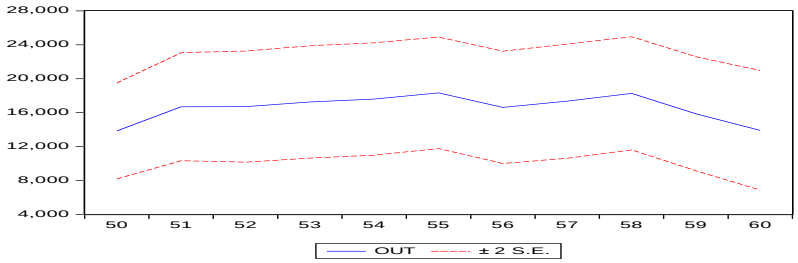
DYNAMIC FORECASTING



Forecast: IN
Actual: Z
Forecast sample: 1 49
Adjusted sample: 14 49
Included observations: 36

Root Mean Squared Error	1475.296
Mean Absolute Error	1056.588
Mean Abs. Percent Error	8.874457
Theil Inequality Coefficient	0.054272
Bias Proportion	0.000011
Variance Proportion	0.022744
Covariance Proportion	0.977245

Figure 1a: In-Sample dynamic forecast

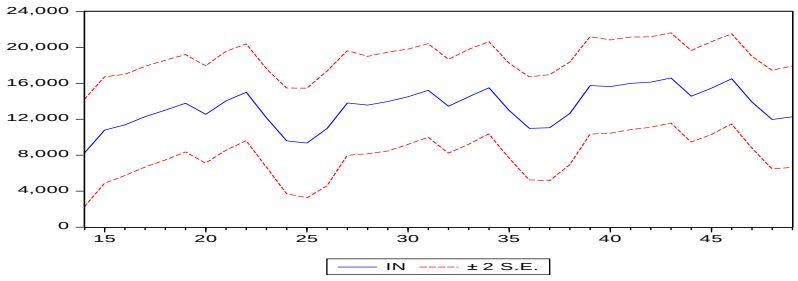


Forecast: OUT
Actual: Z
Forecast sample: 50 60
Included observations: 11

Root Mean Squared Error	1278.318
Mean Absolute Error	1105.357
Mean Abs. Percent Error	6.632363
Theil Inequality Coefficient	0.037946
Bias Proportion	0.049953
Variance Proportion	0.549547
Covariance Proportion	0.400500

Figure 1b: Out-Sample dynamic forecast

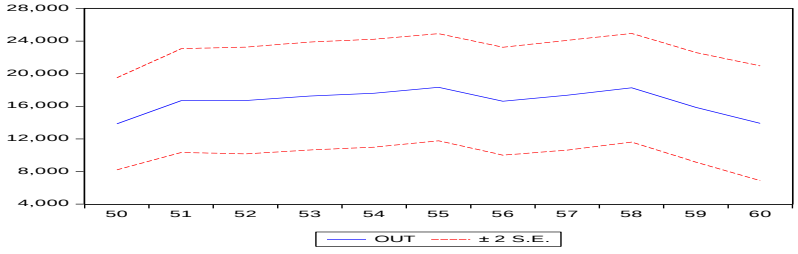
STATIC FORECASTING



Forecast: IN
Actual: Z
Forecast sample: 1 49
Adjusted sample: 14 49
Included observations: 36

Root Mean Squared Error	1412.891
Mean Absolute Error	1038.370
Mean Abs. Percent Error	8.561466
Theil Inequality Coefficient	0.052106
Bias Proportion	0.001190
Variance Proportion	0.069498
Covariance Proportion	0.929312

Figure 2a: In-Sample static forecast



Forecast: OUT
Actual: Z
Forecast sample: 50 60
Included observations: 11

Root Mean Squared Error	1278.318
Mean Absolute Error	1105.357
Mean Abs. Percent Error	6.632363
Theil Inequality Coefficient	0.037946
Bias Proportion	0.049953
Variance Proportion	0.549547
Covariance Proportion	0.400500

Figure 2b: Out-Sample static forecast

REFERENCES

Akaike, H., (1969): Fitting Autoregressive Models for prediction. *Annals of the institute of statistical mathematics*, 21, 243-247

Akaike, H., (1973): Information theory as an extension of the Maximum likelihood principle. *Secondinternational Symposium on information theory*, pp: 265-282

Akaike, H.,(1974): A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716-723

Aras, H., and Aras, N., (2004): Forecasting Residential Natural Gas Demand. *Energy Sources* 26:463-472.

Barron, D. N., (1992): The Analysis of Count Data: Over dispersionand Autocorrelation, Social Methodology, 22, pp 179-220.

Bergstrom, A.R., (1990): Continuous-Time Econometric Modeling.*Oxford University Press*.

Bowerman, B. L.,& Connell O., R.T., (1993): Forecasting and Time Series: *An applied approach.* 3rd Edition.

Bollerslev, T.,(1986): Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31:307-27

Box, G.E.P., & Jenkins, G. M., (1976): Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco.

Box, G.E.P., and Jenkins, G.C., (1976): Time Series Analysis, Fore-casting and Control. *Revised edition.* Holden-Day.

Box, G.E.P., and Jenkins, G.M., (1970):*Time Series Analysis, Forecasting, and Control*. Holden-Day.

Box,G.E.P., Jenkins,G.M., and Reinsel, G.C.,(1994): *Time Series Analysis. Forecasting and Control*,*3rd Ed.* Prentice Hall

Brockwell, P. J., and Davis, R. A., (2002):*Introduction to Time series and forecasting,* 2nd Ed. Springer-Verlag New York, Inc., 170-172

Canova, F., and Hansen, B.E., (1995): "Are Seasonal Patterns Constant over Time? A Test for Seasonal Stability", Journal of Business and Economic Statistics 13(3):236-250.

Chatfield, C., (2004): *The Analysis of Time Series: An Introduction*. Chapman and Hall/CRC

Conejo A. J., Miguel A., Plazas, Rosa E., and Ana B. M., (2005): “Day-ahead electricity price forecasting using the wavelet transform and ARIMA models,” *IEEE Trans. Power Syst.,* vol. 20, no. 2, pp. 1034–1040.

Contreras, J., Rosario E., Francisco J., Nogales, and Conejo A. J., (2003): "ARIMA models to predict next-day electricity prices. Power Systems," *IEEE Transactions on* 2003, vol.18, no. 3, pp. 1012-1018

Crompton, P., and Wu, Y., (2005): Energy consumption in China: past trends and future directions.*Energy Economics* 27: 195-208.

Cryer, J.D. (1986): *Time Series Analysis*. PWS-KENT Publishing Company. Boston

De Livera, A.M., Hyndman, R.J., & Snyder, R. D., (2011): Forecasting time series with complex seasonal patterns using exponential smoothing, Journal of the American Statistical Association, 106(496), 1512-1526.

Dickey, D.A., and Fuller, W.A., (1979): "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", Journal of the American Statistical Association 74:426-431

Doguwa, S. I.S., and Alade, S.O., (2015): On Time Series Modeling of Nigeria's External Reserves; *CBN jounalof Applied Statistics* vol. 6 No. 1(a)

Ediger, V., and Akar, S.,(2007): ARIMA forecasting of primary energy demand by fuel in Turkey.*Energy Policy* 35:1700-1705.

Ekhosuehi, N., (2013): Effects of Box-Cox Transformation on K-th Moving Average Processes for Time Series Forecasting Models. *Journal of the Nigerian Statistical Association* 25: 1-7

Erdoğdu, E. Natural gas demand in Turkey Applied Energy (2010): 87: 211-219.

Fuller, W.A.,(1996): *Introduction to Statistical Time Series, 2nd. Ed*. Wiley Series in Probability and Statistics. John Wiley and Sons, Inc.

Ghiassi, M., Zimbra, D. K.,&Saidane, H., (2006): Medium term system load forecasting with adynamic artificial neural network model. *Electric Power Systems Research* 76(5), PP. 300-314.

Göncü, A., Karahan, M.O., and Kuzubas, T.U. (2011): Pricing of Temperature-based Weather Options for Turkey. *IktisatIsletme ve Finans* 26(309) 33-50

Göncü, A., (2011): Pricing Temperature-based Weather Derivatives in China. *Journal of Risk Finance* 13(1):32–44.

Guerrero, V.M., (1993): Time-series analysis supported by power transformations. Journal of Forecasting, 12, 38–49

Gümrah, F., Katircioglu, D., Aykan, Y., Okumuş, S., and Kilinçer, N. (2001): Modeling of Gas demand using degree-dayconcept: case study for Ankara. *Energy Sources* 23:101-114.

Gupta, S. C., (2007): Fundamentals of Statistics

Hamilton, J. D.,(1994): *Time Series Analysis*. Princeton University Press

Hannan, E.J., and Quinn, B.G., (1979): The determination of the order of an autoregression. *Journal of the Royal Statistical Society, Series B (Methodological)*, 41(2):190-195

Hannan, E.J., (1980): The estimation of the order of an arma process. *Annals of Statistics*, (8):1071-1081

Hannan, E.J., and Rissanen, J., (1982): Recursive estimation of mixed autoregressive-moving average order. *Biometrica*, 69(1):81-94

Hurvich, C. M., and Tsai, C., (1989): Regression and Time series Model selection in small samples.*Biometrica*, 76:290-305

Hyndman, R.J., and Khandakar, Y., (2008): "Automatic time series forecasting: The forecast package for R", Journal of Statistical Software, 26(3).

Jones, R.H. (1975): Fitting autoregressions. *Journal of American Statistics Association*,(70):590-592,

Jones, M.C.,(1987): Randomly choosing parameters from the stationary and invertibility region of autoregressive-moving average models. *Journal of Applied Statistics*,36(2):134-138

Kumar, U., Jain, V. K., (2010): Time series models (Grey-Markov, Grey Model with rolling mechanism and singular spectrum analysis) to forecast consumption in India. *Energy* 35:1709-1716.

Kwiatkowski, D., Phillips, P.C.B., Schmidt P and Shin Y (1992): "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root", Journal of Econometrics 54:158-177.

Kavalieris, L.,(1991): A note on estimating autoregressive-moving average order. *Biometrika*, 78(4):920-922

Linhart, H. and Zucchini, W. (1986): *Model Selection*. John Wiley & Sons

Liu, L.M., and Lin, M.W., (1991): Forecasting residential consumption of natural gas using monthly and quarterly time series. *International Journal of Forecasting* 7:3-16.

Ljung, G. M., & Box, G.E.P., (1978): On a Measure of Lack of fit in Time Series Models. *Biometrika*, Vol. 65(2), pp.297-303

Makridakis, S., Andersen, A., Carbone (1984): The Forecasting Accuracy of Major Time Series Methods, Wiley.

Monahan, J.F.,(1984): A note on enforcing stationarity in autoregressive-moving average models. *Biometrika*, 71:403-404

Musundi, S., Wabomba, M'mukiira P., Mutwiri, (2016): Modeling and Forecasting Kenyan GDP Using Autoregressive Integrated Moving Average (ARIMA) Models:*Science Journal of Applied Mathematics and Statistics*. Vol.4, No. 2, pp. 62-73.

Osborn, D.R., Chui, A.P.L., Smith J, and Birchenhall C.R., (1988): "Seasonality and the order of integration for consumption", Oxford Bulletin of Economics and Statistics 50(4):360-378.

Osborn, D.R. (1990): "A survey of seasonality in UK macroeconomic variables", International Journal of Forecasting, 6:325-337.

Peena D., Tiao, and Tsay,R.S., editors (2001): *A Course in Time Series Analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York

Peiris, M., & Perera, B., (1988): On prediction with fractionally differenced ARIMA models, Journal of Time Series Analysis, 9(3), 213-221.

Peiris, M.S.,(2003): Improving the Quality of Forecasting using Generalized AR Models: An Application to Statistical Quality Control. *Statistical Methods*, 5(2):156-171

Piccolo, D., (1982): The size of the stationary and invertibility region of an autoregressive-moving average process. *Journal of Time Series Analysis*, 3(4):245-247

Rao, T.S.,(1993): editor. *Development in Time Series Analysis*. Chapman & Hal

Saab, Samer, Elie Badr and George Nasr (2001): "Univariate modeling and forecasting of energy consumption: *the case of electricity in Lebanon*," *Energy*, vol.26, pp. 1-12.

Sarak, H., and Satman, A., (2003): The degree-day method to estimate the residential heating natural gas consumption in Turkey: a case study. *Energy* 28: 929-939.

Said, E. and Dickey, D.A., (1984): "Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order", Biometrika 71:598-608.

Schwarz, C., (1978): Estimating the dimension of a model. The Annals of Statistics, 6 (2): 460-465

Sen, L.K., and Shitan, M.,(2002): The performance of aicc as an order selection Criteria in arma time series models. *Pertanika Journal of Science and Technology*, 10(1):25-33

Shenstone, L., and Hyndman, R.J., (2005): "Stochastic models underlying Croston's method for intermittent demand forecasting". Journal of Forecasting, 24, 388-400.

Shibata, R.,(1976): Selection of the order of an autoregressive model by akaike's information Criteria. *Biometrika*, 63(1):117-126

Soldo, B., (2012): Forecasting natural gas consumption. *Applied Energy* 93:26-37.

Sugiura, N.,(1978): Further analysis of the data by Akaike's Information Criteria and the finite corrections. *Communications in Statistics, Theory and Methods*, A7:13-26

Zhu Qing, Yujing Guo, Genfu Feng (2012): "Household energy consumption in China forecasting with BVAR model up to 2015," *2012 Fifth International Joint Conference on Computational Sciences and Optimization*,.